SPRINGER NATURE
Reference

Chittaranjan Kole
Editor

Compendium of Crop Genome Designing for Nutraceuticals

Springer

Compendium of Crop Genome Designing for Nutraceuticals

Chittaranjan Kole
Editor

Compendium of Crop Genome Designing for Nutraceuticals

With 153 Figures and 123 Tables

Springer

Editor
Chittaranjan Kole
Prof. Chittaranjan Kole Foundation for Science and Society
Kolkata, India

ISBN 978-981-19-4168-9 ISBN 978-981-19-4169-6 (eBook)
https://doi.org/10.1007/978-981-19-4169-6

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Paper in this product is recyclable.

Dedication

Dedicated to
My parents, Late Bibhuti Bhushan Kole and Late Bhanumati Kole

Acknowledgement

Special Acknowledgement to
Phullara Kole
For her outstanding assistance in editing this compendium

Preface

The crop plants cater not only to our basic F5 (food, feed, fiber, fuel, and furniture) needs but also provide a number of nutraceuticals with potential nutritional, safety, and therapeutic attributes. Many crop plants provide an array of minerals, for example, Ca, Mg, K, Na, P, Cr, Cu, Fe, Mn, Mo, Ni, Se, and Zn; vitamins, for example, A, B-series, C, E and K; and antioxidant-rich bioactive phytochemicals, including Carotenoids, Phytosterols, Limonoids, Polyphenols, Glucosinolates, phytoestrogen, Terpenoids, Fibers, Polysachharides, and Saponins. Increasing incidences of chronic diseases such as cancer, diabetes and HIV, and malnutrition necessitate global attention to health and nutrition security with equal emphasis on food security. Stupendous amount of researches on biochemical, physiological, and genetic mechanisms underlying the biosynthesis of the health and nutrition related nutraceuticals and precise breeding strategies for augmentation of their content and amelioration of their quality in crop plants are underway all over the world under all commodity categories of crops including cereals and millets, oilseeds, grain legumes, fruits and nuts, and vegetables. This major review work entitled, "Compendium of Crop Genome Designing for Nutraceuticals," comprises five sections dedicated to these five commodity groups and presents enumeration on the concepts, strategies, tools, and techniques of nutraceutomics. These sections include 50 chapters devoted to even number of major crop plants. These chapters present deliberations on the biochemistry and medicinal properties of the nutraceuticals contained; genetic variation of their contents; classical genetics of and breeding for their quantitative and qualitative improvement; tissue culture and genetic engineering for augmentation of productivity and quality; and sources of genes underlying their biosynthesis. They also include comprehensive enumeration on genetic mapping of the genes and QTLs controlling the contents and profile of the nutraceuticals and molecular breeding for their further improvement through marker assisted selection and backcross breeding tools. Prospects of post-genomic precise and targeted breeding strategies including genome-wide association mapping, genomic selection, allele mining and genome editing are also discussed. This compendium would fill the gap in academia, and research and development wings of the private sector industries and will also facilitate understanding of the policy making agencies and people in the socio-economic domain, and benefit students, teachers, scientists, policy makers and sponsoring agencies involved in an array of subjects relevant to

crop sciences, specifically genetics, genomics, tissue culture, genetic engineering, molecular breeding, genomics-assisted breeding, gene editing, bioinformatics, biochemistry, physiology, pathology, entomology, pharmacognosy, and IPR.

I express my thanks to the authors of the chapters of this compendium for their useful contributions and sincere cooperation. I am also grateful to the staff of the publisher for their assistance since inception till completion of editing this compendium.

Kolkata, India
November 2023

Chittaranjan Kole

Contents

Volume 1

Volume 2

About the Editor

Prof. Chittaranjan Kole is an internationally renowned academician with a spectacular professional career of about 40 years!. His pioneering scientific contributions, specifically in the fields of plant genomics and biotechnology, have been globally appreciated. Prof. Kole has developed a number of original concepts and strategies, which have contributed enormously to science and benefited the national as well as global society. His scholarly publications include 150-plus research articles and over 180 books with globally-reputed publishers. Prof. Kole's academic contributions have been profusely appreciated by seven Nobel Laureates including Profs. Norman E. Borlaug, Arthur Kornberg, Werner Arber, Phillip A. Sharp, Günter Blobel, Leland H. Hartwell, and Roger D. Kornberg. His scientific achievements have been honored with several awards, fellowships and recognitions including the "Outstanding Crop Scientist Award" conferred by the International Crop Science Society in recognition of his "life-time and original contributions in the field of crop science."

Prof. Kole worked as a researcher, faculty member and administrator in a large number of premier institutions and universities in India and abroad. In India, he worked across all academic positions from an Assistant Professor to Vice Chancellor in three premier universities including Orissa University of Agriculture and Technology, Sam Higginbottom University of Agriculture, Technology and Sciences, and Bidhan Chandra Krishi Viswavidyalaya. He also worked in Indo-Russian Center for Biotechnology as its First

Project Coordinator. In abroad, he worked in the USSR Academy of Sciences, erstwhile USSR, as a Post-Doctorate Scholar; University of Wisconsin, USA as an Overseas Research Associate; The Pennsylvania State University, USA, and Clemson University, USA, as a Visiting Professor; and Institute of Nutraceutical Research at Clemson University as Director of Research.

Prof. Kole is recognized as a visionary science leader in the global arena. He is the Founding President of three international organizations including the Genome India International, International Climate Resilient Crop Genomics Consortium, and International Phytomedomics and Nutriomics Consortium. In recognition of his international leadership quality, the Food and Agriculture Organization invited Prof. Kole to act as the Leader of the Climate Change theme for the FAO International Symposium on "The Role of Agricultural Biotechnologies in Sustainable Food Systems and Nutrition" in 2016. He organized and chaired many prestigious international workshops, chaired several technical sessions; and delivered innumerable invited plenary lectures and keynote addresses in many international scientific meetings.

Prof. M. S. Swaminathan, World Food Prize Laureate, once wrote to Prof. Kole that "You are a role model for all of us." while Nobel Laureate in Chemistry Prof. Roger D. Kornberg wrote to the Honorable Prime Minister of India about Prof. Kole that "your country will be increasingly benefitted by utilizing his comprehensive knowledge and visionary ideas on science, education and agriculture." Above all, Nobel Laureate in Peace, Dr. Norman E. Borlaug, the Father of Green Revolution, wrote to Prof. Kole that "May all Ph.D.s, future scientists and students that are devoted to agriculture get an inspiration as it refers to your work."

Contributors

Thais A. R. Ramos Universidade Federal da Paraíba, João Pessoa, Brazil

Aakanksha Department of Genetics, University of Delhi South Campus, New Delhi, India

Mohd Amin Abd Halim Malaysian Palm Oil Board, Persiaran Institusi, Bandar Baru Bangi, Kajang, Malaysia

Mostafa Abdelrahman Aswan University Faculty of Science, Aswan, Egypt

Dinesh Adhikary Department of Agricultural, Food & Nutritional Sciences, University of Alberta, Edmonton, AB, Canada

Kumari Aditi Kansas State University, Manhattan, KS, USA

J. Ali Professor Jayashankar Telangana State Agricultural University, Hyderabad, India

Jameel M. Al-Khayri Department of Agricultural Biotechnology, College of Agriculture and Food Sciences, King Faisal University, Al-Ahsa, Saudi Arabia

Victor Acheampong Amankwaah Biotechnology, Seed and Post-Harvest Division, Crops Research Institute, Kumasi, Ghana

R. Ananthan National Institute of Nutrition, Hyderabad, India

J. Aravind Kumar ICAR-Indian Institute of Rice Research, Hyderabad, India

M. A. Ashrutha Hytech seeds India Pvt. Ltd, Hyderabad, India

Shruthi H B Pigeonpea Improvement Program, International Crops Research Institute for Semi-Arid Tropics (ICRISAT), Patancheru, Telangana, India

B. Kalyan Babu ICAR-Indian Institute of Oil Palm Research, Pedavegi, Andhra Pradesh, India

Rachana Baguda Professor Jayashankar Telangana State Agricultural University, Hyderabad, India

M. Balram International Rice Research Institute, Los Banos, Philippines

Kalyana Babu Banisetti ICAR-Indian Institute of Oil Palm Research, Eluru, India

Paramananda Barman Inclusive Health and Traditional Knowledge Studies Division, CSIR-National Institute of Science Communication and Policy Research, New Delhi, India

Surendra Barpete Scientist, Food Legumes Research Platform (FLRP), International Centre for Agricultural Research in the Dry Areas (ICARDA), Sehore, Madhya Pradesh, India

V. Becerra INIA - Instituto de Investigaciones Agropecuarias, Chillán, Chile

Sandip Bera ICAR-Directorate of Groundnut Research, Junagadh, Gujarat, India

N. Bhanushree Division of Vegetable Science, ICAR–Indian Agricultural Research Institute, New Delhi, India

Varun Bhardwaj Department of Botany, University of Delhi, Delhi, India

Jyotsna Bharti Nutritional Improvement of Crops (NIC) Group, International Centre for Genetic Engineering and Biotechnology (ICGGEB), New Delhi, India

Ranjana Bhattacharjee International Institute of Tropical Agriculture, Ibadan, Nigeria

Revanappa Biradar ICAR-Indian Institute of Pulses Research, Regional Station, Dharwad, Karnataka, India

M. W. Blair TSU – Tennessee State University, Department of Agricultural and Environmental Sciences, Nashville, TN, USA

Naresh Bomma Pigeonpea Improvement Program, International Crops Research Institute for Semi-Arid Tropics (ICRISAT), Patancheru, Telangana, India

Alejandro Calle Plant and Environmental Sciences, Clemson University, Clemson, SC, USA

Hugo Campos International Potato Center, Lima, Peru

Hengchun Cao Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, Henan, China

Henan Key Laboratory of Specific Oilseed Crops Genomics (Henan Sesame Research Center, Henan Academy of Agricultural Sciences), Zhengzhou, Henan, China

Henan International Joint Laboratory of Specific Oilseed Crops Improvement, Zhengzhou, Henan, China

Paola Carnevali Barilla S.p.A, Parma, PR, Italy

Domenico Carputo Department of Agricultural Sciences, University of Naples Federico II, Portici, Italy

Ilaria Carrara Dipartimento di Scienze degli Alimenti e del Farmaco, Università degli Studi di Parma, Parma, PR, Italy

A. Carrasco-Pancorbo Department of Analytical Chemistry, Faculty of Sciences, University of Granada, Granada, Spain

Pablo F. Cavagnaro National Scientific and Technical Research Council (CONICET), National Institute of Agricultural Technology (INTA) E.E.A. La Consulta, La Consulta CC8, San Carlos, Mendoza, Argentina

S. Antony Ceasar Division of Plant Molecular Biology and Biotechnology, Department of Biosciences, Rajagiri College of Social Sciences, Kerala, Kochi, India

Ibrahim Çelik Department of Agricultural and Livestock Production, Çal Vocational School of Higher Education, Pamukkale University, Denizli, Turkey

Dipankar Chakraborti Department of Genetics, University of Calcutta, Kolkata, West Bengal, India

Puja Chakraborty Nutritional Improvement of Crops (NIC) Group, International Centre for Genetic Engineering and Biotechnology (ICGGEB), New Delhi, India

Anup Chandra ICAR-Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Prathyusha Cheguri Pigeonpea Improvement Program, International Crops Research Institute for Semi-Arid Tropics (ICRISAT), Patancheru, Telangana, India

Charles Chen Department of Crop, Soil, & Environmental Sciences, Auburn University, Auburn, AL, USA

Vasugi Chinnaiyan Division of Fruit crops, ICAR–Indian Institute of Horticultural Research, Bengaluru, India

Walter Chitarra Council for Agricultural Research and Economics (CREA) – Research Centre for Viticulture and Enology, Conegliano, Italy

Institute for Sustainable Plant Protection, CNR, Torino, Italy

Vishal Chugh College of Horticulture, Banda University of Agriculture and Technology, Banda, India

Isadora Louise Alves da Costa Ribeiro Quintans Universidade Federal Rural do Semi-Árido, Mossoro, Brazil

Ana Paula da Mata Department of Agronomy, Federal University of Viçosa, Viçosa, Brazil

Carla Cristina da Silva Department of Agronomy, Federal University of Viçosa, Viçosa, Brazil

Vincenzo D'Amelia Institute of Biosciences and Bioresources, National Research Council of Italy (CNR-IBBR), Portici, Italy

Vijayalaxmi S. Dandin Department of Biology, JSS College, Dharwad, India

Alok Das Division of Plant Biotechnology, ICAR-Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Arpita Das Department of Genetics & Plant Breeding, Bidhan Chandra Krishi Viswavidyalaya, Mohanpur, Nadia, West Bengal, India

Stefano Delbono Council for Agricultural Research and Economics (CREA), Research centre for Genomics & Bioinformatics, Fiorenzuolad'Arda, PC, Italy

G. P. Dixit ICAR-Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Ivica Djalovic Institute of Field and Vegetable Crops, National Institute of the Republic of Serbia, Novi Sad, Serbia

Teresa Docimo Institute of Biosciences and Bioresources, National Research Council of Italy (CNR-IBBR), Portici, Italy

Sami Doğanlar Plant Science and Technology Application and Research Center, Izmir Institute of Technology, Urla, Turkey

Department of Molecular Biology and Genetics, Izmir Institute of Technology, Urla, Turkey

Yinghui Duan Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, Henan, China

Henan Key Laboratory of Specific Oilseed Crops Genomics (Henan Sesame Research Center, Henan Academy of Agricultural Sciences), Zhengzhou, Henan, China

Henan International Joint Laboratory of Specific Oilseed Crops Improvement, Zhengzhou, Henan, China

Anil Kumar Dubey Division of Fruits and Horticultural Technology, ICAR-Indian Agricultural Research Institute, New Delhi, India

Frank Dunemann Julius Kühn-Institut (JKI), Federal Research Centre for Cultivated Plants, Institute for Breeding Research on Horticultural Crops, Quedlinburg, Germany

Sudip Kumar Dutta Gus R. Douglass Institute, Department of Biology, West Virginia State University, Institute, WV, USA

ICAR RC NEH Region, Sikkim centre, Gangtok, Sikkim, India

Maria Stefanie Dwiyanti Laboratory of Applied Plant Genome, Hokkaido University, Sapporo, Japan

Murugesh Easwaran Nutritional Improvement of Crops (NIC) Group, International Centre for Genetic Engineering and Biotechnology (ICGGEB), New Delhi, India

L. Elkonin Department of Biotechnology, Federal Centre of Agriculture Research of the South-East Region, Saratov, Russia

Magdi El-sayed Faculty of Science, Galala University, Suze, Egypt

Jun Fan Yunnan Academy of Agricultural Sciences, Kunming, China

Khaled Fathy Nutritional Improvement of Crops (NIC) Group, International Centre for Genetic Engineering and Biotechnology (ICGGEB), New Delhi, India

Yolanda Ferradás Instituto de Ciencias de la Vid y del Vino (ICVV), Consejo Superior de Investigaciones Científicas– Universidad de la Rioja – Gobierno de La Rioja, Logroño, Spain

Faculty of Biology, University of Santiago de Compostela, Santiago de Compostela, Spain

Franca Finocchiaro Council for Agricultural Research and Economics (CREA), Research centre for Genomics & Bioinformatics, Fiorenzuolad'Arda, PC, Italy

Anne Frary Department of Molecular Biology and Genetics, Izmir Institute of Technology, Urla, Turkey

Prakash I. Gangashetty Pigeonpea Improvement Program, International Crops Research Institute for Semi-Arid Tropics (ICRISAT), Patancheru, Telangana, India

Marica Gasparro Council for Agricultural Research and Economics (CREA) – Research Centre for Viticulture and Enology, Turi, Italy

Thais Gaudencio do Rêgo Universidade Federal da Paraíba, João Pessoa, Brazil

Roberta Ghizzoni Council for Agricultural Research and Economics (CREA), Research centre for Genomics & Bioinformatics, Fiorenzuolad'Arda, PC, Italy

Sanatan Ghosh Department of Genetics, University of Calcutta, Kolkata, West Bengal, India

Shailendra Goel Department of Botany, University of Delhi, Delhi, India

J. J. Gonzalez-Fernandez Instituto de Hortofruticultura Subtropical y Mediterranea La Mayora (IHSM La Mayora-UMA-CSIC), Malaga, Spain

Poulami Goswami Department of Botany, University of Delhi, Delhi, India

Velu Govindan International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico

Mahalingam Govindaraj HarvestPlus Program, Alliance of Bioversity International and the International Center for Tropical Agriculture (CIAT), Cali, Colombia

Maria D. P. T. Gunawan-Puteri Department of Food Technology, Swiss German University, Tangerang, Indonesia

DebJyoti Sen Gupta ICAR-Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Mehak Gupta Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, India

Om Prakash Gupta ICAR-Indian Institute of Wheat and Barley Research, Karnal, India

Sanjeev Gupta Assistant Director General, Oil seeds & Pulses, Indian Council of Agricultural Research, New Delhi, India

Nergiz Gürbüz Çolak Plant Science and Technology Application and Research Center, Izmir Institute of Technology, Urla, Turkey

S. M. Harisha Division of Vegetable Science, ICAR–Indian Agricultural Research Institute, New Delhi, India

Anjan Hazra Department of Genetics, University of Calcutta, Kolkata, West Bengal, India

J. I. Hormaza Instituto de Hortofruticultura Subtropical y Mediterranea La Mayora (IHSM La Mayora-UMA-CSIC), Malaga, Spain

Massimo Iorizzo Plants for Human Health Institute and Department of Horticultural Sciences, North Carolina State University, Kannapolis, NC, USA

Maizura Ithnin Malaysian Palm Oil Board, Persiaran Institusi, Bandar Baru Bangi, Kajang, Malaysia

Rani Asaram Jadhav College of Agriculture, Nagpur, India

Arun Jagannath Department of Botany, University of Delhi, Delhi, India

S. Mohan Jain Department of Agricultural Sciences, University of Helsinki, Helsinki, Finland

Souframanien Jegadeesan Nuclear Agriculture & Biotechnology Division, Bhabha Atomic Research Centre, Trombay, Mumbai, India

Dennis V. Johnson Agriculture Consultant, Middlebrook Ave, Cincinnati, OH, USA

Ming Ju Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, Henan, China

Henan Key Laboratory of Specific Oilseed Crops Genomics (Henan Sesame Research Center, Henan Academy of Agricultural Sciences), Zhengzhou, Henan, China

Henan International Joint Laboratory of Specific Oilseed Crops Improvement, Zhengzhou, Henan, China

Sai Rekha Kadirimangalam International Crop Research Institute for the Semi-Arid Tropics, Hyderabad, India

Pritam Kalia Division of Vegetable Science, ICAR-Indian Agricultural Research Institute, New Delhi, India

Anmol Kalra Department of Botany, University of Delhi, Delhi, India

Rashmi Kaul Nutritional Improvement of Crops (NIC) Group, International Centre for Genetic Engineering and Biotechnology (ICGGEB), New Delhi, India

Tanushri Kaul Nutritional Improvement of Crops (NIC) Group, International Centre for Genetic Engineering and Biotechnology (ICGGEB), New Delhi, India

Gurpreet Kaur Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, India

Prashant Kaushik Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Valencia, Spain

Mercy Ketavi Research Technology Support Facility, Michigan State University, East Lansing, MI, USA

Jasminkumar Kheni Department of Biotechnology, Junagadh Agricultural University, Junagadh, Gujarat, India

Rituraj Khound Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Scottsbluff, NE, USA

Chittaranjan Kole Institute of Nutraceutical Research, Clemson University, Clemson, SC, USA

Aravind Konda ICAR-Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Schuyler S. Korban Department of Natural Resources and Environmental Sciences, University of Illinois at Urbana–Champaign, Urbana, IL, USA

Shivanand Koti Department of Fruit Science College of Horticulture, UHSB Campus, Bangalore, India

Ajay Kumar ICAR-Indian Institute of Horticultural Research, Bengaluru, India

Jitendra Kumar ICAR-Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Sunil Kumar ICAR-Indian Institute of Wheat and Barley Research, Karnal, India

Rituparna Kundu Chaudhuri Department of Botany, Barasat Government College, Barasat, West Bengal, India

M. T. Labuschagne Department of Plant Sciences, University of the Free State, Bloemfontein, South Africa

Amrit Lamichaney ICAR-Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Charu Lata Inclusive Health and Traditional Knowledge Studies Division, CSIR-National Institute of Science Communication and Policy Research, New Delhi, India

Chun Li Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, Henan, China

Henan Key Laboratory of Specific Oilseed Crops Genomics (Henan Sesame Research Center, Henan Academy of Agricultural Sciences), Zhengzhou, Henan, China

Henan International Joint Laboratory of Specific Oilseed Crops Improvement, Zhengzhou, Henan, China

H. Li TSU – Tennessee State University, Department of Agricultural and Environmental Sciences, Nashville, TN, USA

Mathiazhagan Malarvizhi Division of Biotechnology, ICAR Indian Institute of Horticultural Research, Bengaluru, India

S. Malathi ICAR-Indian Institute of Rice Research, Hyderabad, India

Manisha Mangal Division of Vegetable Science, ICAR-Indian Agricultural Research Institute, New Delhi, India

S. K. Mangrauthia ICAR-Indian Institute of Rice Research, Hyderabad, India

C. Manimurugan ICAR-Indian Institute of Oilseeds Research, Hyderabad, India

B. Manu ICAR-Indian Institute of Pulses Research, Regional Station, Dharwad, Karnataka, India

Malarvizhi Mathiazhagan Division of Biotechnology, ICAR–Indian Institute of Horticultural Research, Bengaluru, India

Ramesh Kanna Mathivanan Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Scottsbluff, NE, USA

Shikha Mathur Department of Genetics, University of Delhi South Campus, New Delhi, India

Tesfaye Walle Mekonnen Department of Plant Sciences, University of the Free State, Bloemfontein, South Africa

Hongmei Miao Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, Henan, China

Henan Key Laboratory of Specific Oilseed Crops Genomics (Henan Sesame Research Center, Henan Academy of Agricultural Sciences), Zhengzhou, Henan, China

Henan International Joint Laboratory of Specific Oilseed Crops Improvement, Zhengzhou, Henan, China

Rosa Anna Milella Council for Agricultural Research and Economics (CREA) – Research Centre for Viticulture and Enology, Turi, Italy

Chandra Nath Mishra ICAR-Indian Institute of Wheat and Barley Research, Karnal, India

Loredana Moffa Council for Agricultural Research and Economics (CREA) – Research Centre for Viticulture and Enology, Conegliano, Italy

Suma C. Mogali AICRP-MULLaRP Scheme, University of Agricultural Sciences, Dharwad, Karnataka, India

Bidisha Mondal Department of Genetics and Plant Breeding, The School of Agriculture and Allied Sciences, The Neotia University, Sarisha, West Bengal, India

Biswajit Mondol Division of Plant Biotechnology, ICAR-Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Caterina Morcia Council for Agricultural Research and Economics (CREA), Research centre for Genomics & Bioinformatics, Fiorenzuolad'Arda, PC, Italy

Samwel Muiruri International Institute of Tropical Agriculture (IITA), Nairobi, Kenya

Department of Plant Sciences, Kenyatta University, Nairobi, Kenya

Aloleca Mukherjee Department of Genetics, University of Calcutta, Kolkata, West Bengal, India

Epsita Mukherjee Amity Institute of Biotechnology, Amity University, Noida, India

T. K. Nagarathna University of Agricultural Sciences, GKVK Campus, Bengaluru, Karnataka, India

Praveen Nagella Department of Life Sciences, CHRIST (Deemed to be University), Bangalore, India

Poornananda M. Naik Department of Botany, Karnatak University, Dharwad, Karnataka, India

Aswathy G. H. Nair Division of Crop Improvement, Central Tuber Crops Research Institute, Thiruvananthapuram, India

Manivannan Narayana Centre of Excellence in Molecular Breeding, Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore, India,

Spurthi N. Nayak Department of Biotechnology, University of Agricultural Sciences, Dharwad, India

C. N. Neeraja ICAR-Indian Institute of Rice Research, Hyderabad, India

Mamta Nehra Nutritional Improvement of Crops (NIC) Group, International Centre for Genetic Engineering and Biotechnology (ICGGEB), New Delhi, India

L. Nekkalapudi TSU – Tennessee State University, Department of Agricultural and Environmental Sciences, Nashville, TN, USA

Luca Nerva Council for Agricultural Research and Economics (CREA) – Research Centre for Viticulture and Enology, Conegliano, Italy

Institute for Sustainable Plant Protection, CNR, Torino, Italy

José Ribamar Costa Ferreira Neto Universidade Federal de Pernambuco, Recife, Brazil

G. Niharika ICAR-Indian Institute of Rice Research, Hyderabad, India

Padma Nimmakayala Gus R. Douglass Institute, Department of Biology, West Virginia State University, Institute, WV, USA

Valentine Otang Ntui International Institute of Tropical Agriculture (IITA), Nairobi, Kenya

L. Olmo-García Department of Analytical Chemistry, Faculty of Sciences, University of Granada, Granada, Spain

Bonny Michael Oloka Department of Horticultural Science, North Carolina State University, Raleigh, NC, USA

Abrizah Othman Malaysian Palm Oil Board, Persiaran Institusi, Bandar Baru Bangi, Kajang, Malaysia

Kadirvel Palchamy ICAR-Indian Institute of Oilseeds Research, Hyderabad, Telangana, India

Manish Pandey International Crops Research Institute for the Semiarid Tropics, Hyderabad, Telangana, India

Sudhakar Pandey ICAR-Indian Institute of Vegetable Research, Varanasi, India

Valesca Pandolfi Universidade Federal de Pernambuco, Recife, Brazil

M. Paredes INIA - Instituto de Investigaciones Agropecuarias, Chillán, Chile

Univ Vina del Mar, Escuela de Ciencas Agricolas y veterinarias, Vina del Mar, Chile

A. K. Parihar ICAR-Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Guilherme Silva Pereira Department of Agronomy, Federal University of Viçosa, Viçosa, Brazil

K. Poornima ICAR-Indian Institute of Horticultural Research, Bengaluru, India

Ameena Premnath Department of Plant Biotechnology, Centre for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore, India,

Mahesh Pujar Crop Improvement Theme, Research Program-Asia, at ICRISAT, Hyderabad, India

Dhanasekar Punniyamoorthy Nuclear Agriculture & Biotechnology Division, Bhabha Atomic Research Centre, Trombay, Mumbai, India

Naveen Puppala Agricultural Science Center at Clovis, New Mexico State University, Clovis – New Mexico, USA

Jose Quero-García UMR Biologie du Fruit et Pathologie, INRAE, Univ. Bordeaux, Villenave d'Ornon, France

Rawan Rabie Faculty of Science, Galala University, Suze, Egypt

P. Raghuveer Rao ICAR-Indian Institute of Rice Research, Hyderabad, India

Ashutosh Rai College of Horticulture, Banda University of Agriculture and Technology, Banda, India

Avi Raizada Nuclear Agriculture & Biotechnology Division, Bhabha Atomic Research Centre, Trombay, Mumbai, India

M. K. Rajesh Division of Crop Improvement, ICAR-Central Plantation Crops Research Institute, Kasaragod, India

Nirala Ramchiary School of Life Sciences, Jawaharlal Nehru University, New Delhi, India

Bala Rathinasabapathi Horticultural Sciences Department, University of Florida, Gainesville, FL, USA

K. V. Ravishankar ICAR-Indian Institute of Horticultural Research, Bengaluru, India

Kundapura V. Ravishankar Division of Biotechnology, ICAR–Indian Institute of Horticultural Research, Bengaluru, India

Ramya Ravishankar Sun Valley Family Care, Peoria, AZ, USA

Umesh K. Reddy Gus R. Douglass Institute, Department of Biology, West Virginia State University, Institute, WV, USA

Fabio Reggiani Council for Agricultural Research and Economics (CREA), Research centre for Genomics & Bioinformatics, Fiorenzuolad'Arda, PC, Italy

João Ricardo Bachega Feijó Rosa Department of Agronomy, Federal University of Viçosa, Viçosa, Brazil

Anirban Roy Ramakrishna Mission Vivekananda Educational and Research Institute, Kolkata, West Bengal, India

P. R. Saabale ICAR-Indian Institute of Pulses Research, Regional Station, Dharwad, Karnataka, India

Deepti B. Sagare Bayer Crop Science Ltd, Hyderabad, India

Partha Saha Division of Vegetable Science, ICAR–Indian Agricultural Research Institute, New Delhi, India

P. Sanghamitra ICAR-National Rice Research Institute, Cuttack, India

D. Sanjeeva Rao ICAR-Indian Institute of Rice Research, Hyderabad, India

Iara Gonçalves dos Santos Department of Agronomy, Federal University of Viçosa, Viçosa, Brazil

Dipak K. Santra Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Scottsbluff, NE, USA

Abhimanyu Sarker John Innes Centre, Norwich, UK

Jyothish Madambikattil Sasi Inclusive Health and Traditional Knowledge Studies Division, CSIR-National Institute of Science Communication and Policy Research, New Delhi, India

Stefania Savoi Department of Agricultural, Forest and Food Sciences, University of Turin, Grugliasco, Italy

Nunzia Scotti Institute of Biosciences and Bioresources, National Research Council of Italy (CNR-IBBR), Portici, Italy

Joseph Kadanthottu Sebastian Department of Life Sciences, CHRIST (Deemed to be University), Bangalore, India

Raman Selvakumar Centre for Protected Cultivation Technology, ICAR-Indian Agricultural Research Institute (IARI), New Delhi, India

Debjyoti Sen Gupta ICAR-Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Manuel Joaquín Serradilla Centro de Investigaciones Científicas y Tecnológicas de Extremadura (CICYTEX), Instituto Tecnológico Agroalimentario de Extremadura (INTAEX), Badajoz, Spain

Priya Shah International Crops Research Institute for the Semiarid Tropics, Hyderabad, Telangana, India

Megha Sharma Department of Botany, University of Delhi, Delhi, India

Nimisha Sharma Division of Fruits and Horticultural Technology, ICAR-Indian Agricultural Research Institute, New Delhi, India

Prashant Shetti Banashankari Agro Agency, Nipani, India

Masayoshi Shigyo Laboratory of Vegetable Crop Science, College of Agriculture, Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Yamaguchi, Japan

Philipp W. Simon USDA Agricultural Research Service, Vegetable Crops Research Unit, Department of Horticulture, University of Wisconsin, Madison, WI, USA

Gyanendra Singh ICAR-Indian Institute of Wheat and Barley Research, Karnal, India

Jugpreet Singh LeafWorks Inc., Sebastopol, CA, USA

Neetu Singh ICAR-Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Prateek Singh Division of Plant Biotechnology, ICAR-Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Shrawan Singh Division of Vegetable Science, ICAR-Indian Agricultural Research Institute, New Delhi, India

Shweta Singh Indian Institute of Sugarcane Research, Lucknow, India

Dwaipayan Sinha Department of Botany, Government General Degree College Mohanpur, Mohanpur, West Bengal, India

N. Siromani ICAR-Indian Institute of Rice Research, Hyderabad, India

Sonia Khan Sony Nutritional Improvement of Crops (NIC) Group, International Centre for Genetic Engineering and Biotechnology (ICGGEB), New Delhi, India

S. Gandhadmath Spoorti ICAR-Indian Institute of Rice Research, Hyderabad, India

M. Sujatha ICAR-Indian Institute of Oilseeds Research, Hyderabad, India

R. M. Sundaram ICAR-Indian Institute of Rice Research, Hyderabad, India

Penna Suprasanna Amity Institute of Biotechnology, Amity University of Maharashtra (AUM), Mumbai, Maharashtra, India

Noor Idayu Mhd Tahir Malaysian Palm Oil Board, Persiaran Institusi, Bandar Baru Bangi, Kajang, Malaysia

A. Talavera Instituto de Hortofruticultura Subtropical y Mediterranea La Mayora (IHSM La Mayora-UMA-CSIC), Malaga, Spain

Rachele Tamburino Institute of Biosciences and Bioresources, National Research Council of Italy (CNR-IBBR), Portici, Italy

Javier Tello Instituto de Ciencias de la Vid y del Vino (ICVV), Consejo Superior de Investigaciones Científicas– Universidad de la Rioja – Gobierno de La Rioja, Logroño, Spain

Valeria Terzi Council for Agricultural Research and Economics (CREA), Research centre for Genomics & Bioinformatics, Fiorenzuolad'Arda, PC, Italy

Shallu Thakur Division of Plant Biotechnology, ICAR-Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Arulprakash Thangaraj Nutritional Improvement of Crops (NIC) Group, International Centre for Genetic Engineering and Biotechnology (ICGGEB), New Delhi, India

Bhoopal Singh Tomar Division of Vegetable Science, ICAR–Indian Agricultural Research Institute, New Delhi, India

Rukam S. Tomar Department of Biotechnology, Junagadh Agricultural University, Junagadh, Gujarat, India

Jaindra Nath Tripathi International Institute of Tropical Agriculture (IITA), Nairobi, Kenya

Kuldeep Tripathi Division of Germplasm Evaluation, ICAR-National Bureau of Plant Genetic Resources, New Delhi, India

Leena Tripathi International Institute of Tropical Agriculture (IITA), Nairobi, Kenya

Giorgio Tumino Plant Breeding, Wageningen University and Research, Wageningen, The Netherlands

Jayashree Ugalat Department of Biotechnology and Crop Improvement College of Horticulture, UHSB Campus, Bangalore, India

C. Vasugi ICAR-Indian Institute of Horticultural Research, Bengaluru, India

J. Veerendra ICAR-Indian Institute of Rice Research, Hyderabad, India

Rachana Verma Nutritional Improvement of Crops (NIC) Group, International Centre for Genetic Engineering and Biotechnology (ICGGEB), New Delhi, India

Clizia Villano Department of Agricultural Sciences, University of Naples Federico II, Portici, Italy

Ana Wünsch Centro de Investigación y Tecnología Agroalimentaria de Aragón, Zaragoza, Spain

Instituto Agroalimentario de Aragón-IA2 (CITA-Universidad de Zaragoza), Zaragoza, Spain

Fan Xingming Yunnan Academy of Agricultural Sciences, Kunming, China

Bal Govind Yadav International Centre for Genetic Engineering and Biotechnology, New Delhi, India

Satish Kumar Yadava Centre for Genetic Manipulation of Crop Plants, University of Delhi South Campus, New Delhi, India

Shrikant Yankanchi Indira Gandhi Krishi Vishwavidyalaya, Raipur, India

A. Zanwar Interactive Research School for Health Affairs, Bharati Vidyapeeth (Deemed to be University), Pune, India

Haiyang Zhang Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, Henan, China

Henan Key Laboratory of Specific Oilseed Crops Genomics (Henan Sesame Research Center, Henan Academy of Agricultural Sciences), Zhengzhou, Henan, China

Henan International Joint Laboratory of Specific Oilseed Crops Improvement, Zhengzhou, Henan, China

Part I

Cereal Crops

Redesigning Rice as a Promising Nutraceutical Functional Food

R. M. Sundaram, D. Sanjeeva Rao, P. Sanghamitra, S. Gandhadmath Spoorti, J. Veerendra, N. Siromani, G. Niharika, R. Ananthan, J. Aravind Kumar, P. Raghuveer Rao, S. Malathi, S. K. Mangrauthia, M. Balram, J. Ali, and C. N. Neeraja

Contents

R. M. Sundaram · D. Sanjeeva Rao · S. G. Spoorti · J. Veerendra · N. Siromani · G. Niharika · J. Aravind Kumar · P. Raghuveer Rao · S. Malathi · S. K. Mangrauthia · C. N. Neeraja (✉)
ICAR-Indian Institute of Rice Research, Hyderabad, India

P. Sanghamitra
ICAR-National Rice Research Institute, Cuttack, India

R. Ananthan
National Institute of Nutrition, Hyderabad, India

M. Balram
International Rice Research Institute, Los Banos, Philippines

J. Ali
Professor Jayashankar Telangana State Agricultural University, Hyderabad, India
e-mail: j.ali@irri.org

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_1

Abstract

Rice is a staple food for more than half of the world's population and is grown in more than 100 countries in varied water regimes and ecological systems. In recent years, the perspective of rice has changed from staple food to its potential to be used as a functional food. Industrialization brought changes in lifestyle, leading to the consumption of milled rice, leaving aside the rice bran. Brown rice is a good source of energy, fats, vitamins, and minerals, and the colored rice grains have higher mineral content, antioxidants, and bioactive compounds; the polished rice is a poor source of nutrients. Challenges like climate change with elevated CO_2, and drought and heat stress are also reducing the nutritional quality of rice. Enhancing the nutritive value of rice grain and promoting as a nutraceutical functional food could address nutrition security. Breeding interventions and the application of next-generation technologies can hasten the development of nutritive rice varieties with desired levels of the mineral, vitamin, and bioactive compounds and glycemic index. Landraces and wild species are the potential genetic donors aiding in the generation of the breeding material with the increased functionality of rice grain. The availability of enormous rice germplasm gives scope to identify new nutraceuticals and develop nutraceutical-rich varieties. Genomic regions and genes associated with nutritive function in grain are being identified by deploying sequencing, resequencing, genome-wide association, and biparental mapping. CRISPR-based genome editing appears to be the most potent tools for developing rice varieties with high grain nutrient levels.

Keywords

Rice · Grain · Functional food · Nutritive composition · Bioactive compounds · Germplasm · Biofortification · Next-generation technologies

1 Introduction

1.1 Rice: A Staple Food Across the World

Rice is a major cereal crop and a staple food for half of the world's population (Fukagawa and Ziska 2019). Globally, rice is cultivated in more than 100 countries, spanning an area of approximately 163 million hectares (FAO 2021). Rice has two cultivated species, *Oryza sativa,* which is predominant worldwide, and *Oryza*

glaberrima, limited to a few parts of Western Africa (Khush 1997). Within *O. sativa*, two major subspecies have been identified: *indica*, cultivated in tropical Asia, and *japonica*, primarily found in Southern China, South East Asia, Indonesia, and outside Asia (Londo et al. 2006). Later, five distinct groups were made in *O. sativa* corresponding to *indica*, *aus*, *aromatic*, temperate *japonica,* and tropical *japonica* (Garris et al. 2005).

Rice is cultivated worldwide in varied water regimes, ecology systems, and soil types. Nearly three-quarters world's rice is produced from irrigated lowland systems with two to three crops per year (De Datta 1981). Around 20% of global rice production is from the rainfed lowland system, which is predominant in Southeast Asia and Africa. The upland condition accounts for approximately 4% of the world's rice production. Deepwater, semi-deep water, and floating kinds of rice are cultivated in specific geographical regions in some South Asian countries (Khush 1987).

Worldwide paddy production was around 0.8 billion tons in 2019, with an 8% contribution to global crop production (FAO 2021). China, India, and Indonesia are major rice-producing countries, accounting for 50% of rice production (www.fao.org). Other Asian countries like Bangladesh, Vietnam, Myanmar, Thailand, the Philippines, Japan, Pakistan, Cambodia, the Republic of Korea, Nepal, and Sri Lanka contribute to the remaining world's total rice production. Brazil, the United States, Egypt, Madagascar, and Nigeria produce around 5% of global rice production (Muthayya et al. 2014).

Nearly 490 million tons were reported to be global rice consumption in 2019. Daily rice consumption is also the highest in Bangladesh, the Lao People's Democratic Republic, Cambodia, Vietnam, Myanmar, Thailand, Indonesia, and the Philippines among the Asian countries. Significant rice consumption is also indicated in South America, Latin America, the Caribbean, and Oceania, with China and India accounting for ~50% of the world's rice consumption. Interestingly, only 60% of the consumption in Africa is covered by local rice production, with the remaining rice being imported (www.africarice.org). Rice is reported to be the source of 20% of the world's dietary energy, supplying more than 70% of calories in some Asian countries (GRiSP 2013).

Rice production has significantly increased, more than doubling after the advent of the Green Revolution from yield levels of less than 2 t/ha during the 1950s. The development and broad adoption of semi-dwarf rice varieties responsive to increased inputs of fertilizers, irrigation, pesticides, and other resources considerably enhanced rice yield (Dalrymple 1986; Hedden 2003). Rice production was reported to be increased by 130% between the 1960s and 2000s (Muthayya et al. 2014). The next level of yield enhancement was made possible with the exploitation of heterosis by developing hybrid rice during the 1970s in China and other countries (Cheng et al. 2007a).

1.2 Growing Importance of Rice in the Face of Chronic Diseases and Malnutrition

Rice is the staple diet of more than 3.5 billion people (Xu et al. 2021a), mainly in Asian and African countries populated by people with poor purchasing power and access to nutritious food. The world's population is projected to be 8.5 billion in

2030, 9.7 billion in 2050, and 10.9 billion in 2100 (United Nations Department of Economic and Social Affairs 2019). Most of the population growth is expected in Southern Asia and Africa, which are already excessively dependent on rice. Thus, rice has become a crucial commodity in global food security, and around a 70% increase in food production is needed to meet the food demand in 2050 (FAO 2017). The anticipated increase in rice production could be possible from India, China, Vietnam, and Thailand as per projections by the OECD-FAO Agricultural Outlook, global rice production (OECD-FAO 2021).

In addition to being targeted for the enhancement of its production for the future, rice is also being aimed for its nutraceutical quality improvement. People who subsist on polished rice are vulnerable to vitamin and mineral deficiencies. Increasing the nutritive value of rice grain can address nutrition security and food security, especially in countries where rice is the major source of calories/energy. More than 32% of women worldwide and 36.6% in Asia are reportedly anemic (FAO 2019; https://globalnutritionreport.org/reports/global-nutrition-report-2018/). The Global Nutrient Database of availability of macro- and micronutrients in 195 countries from 1980 to 2013 revealed higher consumption of carbohydrates in South Asia against the world's consumption. It has also reported the poor availability of macro- and micronutrients per day per person in South Asia, viz., 47 g fat (world: 72 g), 56 g protein (world: 71 g), 362 μg vitamin A (world: 705 μg), 19 g iron (Fe) (world: 18 mg), and 8 mg zinc (Zn) (world: 10 mg) (https://nutrition.healthdata.org/global-nutrient-database). Half of the world's malnourished children reside in three countries of South Asia, viz., Bangladesh, India, and Pakistan (World Bank 2009). Concerning the Global Nutrient Database, the values of availability of energy (2500 to <2500 kcal), saturated fat (150 to <250 kcal), and protein (200 < 250 kcal) were much lesser in most of the Asian countries than the recommended levels (Schmidhuber et al. 2018). The Comprehensive National Nutritional Survey (CNNS) of India was conducted during 2016–2018 on 38,060 preschool children (aged 0–4 years), 38,355 school-age children (aged 5–9 years), and 35,830 adolescents (aged 10–19 years). It revealed that in children aged 0–4 years 35% were stunted and 17% wasted; in 5- to 9-year-olds, 22% were stunted and 6% were severely stunted, and 35% were underweight, with 10% severely underweight. Nearly one-quarter of children aged 5–9 years (23%) and adolescents aged 10–19 years (24%) were underweight. Anemia was most severe (>50%) among the children (<2 years), 41% of preschoolers, 24% of school-age children, and 28% of adolescents. A higher incidence of anemia was observed in female adolescents (40%) compared to boys (18%). Prevalence of Fe deficiency was also seen in preschoolers (32%), female adolescents (31%), school-age children (17%), and male adolescents (12%). Nearly 17–19% of children and 32% of adolescents were Zn deficient. Among vitamins, deficiency of vitamin A was ~16–22%, vitamin D was 14–24%, vitamin B_{12} was 14–17% in children and 31% in adolescents, and deficiency of folate was also seen in 23–28% among children and 37% of adolescents.

The enhanced rice yield levels are also vulnerable to the significant decline in grain micronutrient content reported in cereal varieties developed over the last few

decades. The decline could be due to the selective breeding of cultivars with higher yields but lesser micronutrient density because of the dilution effect (Marles 2017; McDonald et al. 2008). Though the production of nutrient-rich nonstaples such as vegetables, pulses, fruits, and animal products has also increased, their affordability is less than staple cereals for people with low purchasing power.

Climate change poses a serious challenge to improving the nutritive status of cereals, including rice, thereby threatening global nutritional security. Excess soil moisture/drought/heat and elevated CO_2 could lead to a reduction in nutritional quality in terms of protein, micronutrients, and vitamins in wheat and rice (Soares et al. 2019; Soba et al. 2019; Zhu et al. 2018) and also decrease the concentration of protein in grains of many crops species (Smith and Myers 2018), thereby directly affecting human nutrition (Toreti et al. 2019). As per the study, elevated CO_2 could lead to deficiencies of Zn for 175 million people, protein for 122 million, and Fe for 1.4 billion women of childbearing age and young children (Smith and Myers 2019).

Staple cereals have become major food source during the lockdown period of the COVID-19 pandemic in several Asian countries because of their affordability by low-income groups and ready availability, unlike perishable food items (Bairagi et al. 2022; Neeraja et al. 2022). Despite the disproportionate hikes in noncereal food prices, cereals like rice and wheat prices were stable during and after COVID-19 lockdown (Rathod et al. 2022). Any incremental improvement in the nutritional status of staple cereals would directly impact addressing the nutrient/micronutrient malnutrition of the world. Appropriate agricultural interventions help enrich micronutrient density and encourage dietary diversity could be ideal and long-term sustainable strategies (Bouis et al. 2019). Biofortification could be a suitable agricultural intervention to enrich micronutrient density in commonly consumed cereals such as rice, wheat, and other cereals. Biofortification refers to the genetic enhancement of crucial food crops with enhanced nutrients through agronomy, breeding, and biotechnology strategies (Bouis and Saltzman 2017).

Biofortification is different from fortification, wherein nutrients are externally added to the food items. Since global rice production is concentrated in a few Asian countries, fortification of rice for local consumption and export is proposed to be a strategy to address vitamin and mineral deficiency (Muthayya et al. 2014). Biofortification can be integral to achieving the third Sustainable Development Goal (SDG), focusing on good health and well-being.

1.3 Development of Biofortified Rice Varieties: Limitations of Conventional Breeding and Rationale for Next-Generation Breeding

The development of high-yielding cereals during the Green Revolution era helped many countries in Asia and globally to address food security. However, it led to excessive dependence on cereals as a food source, the rise of micronutrient malnutrition, or hidden hunger, which has been observed globally (Bouis et al. 2019). The initiatives of the Consultative Group on International Agricultural Research

(CGIAR) during the late 1990s and early 2000s led to the HarvestPlus Challenge Program across the world aiming to develop biofortified staple food crops through conventional plant breeding (Sanjeeva Rao et al. 2020).

Crop improvement requires genetic variability within and between the species, and this variability can be assessed by screening with different molecular or genetic markers, which are specific DNA sequences with known locations on chromosomes. They are classified into two categories: classical (morphological, cytological, and biochemical markers) and DNA markers. DNA markers help in the precise characterization of germplasm, construction of linkage maps, and DNA fingerprinting of crop varieties that ultimately result in marker-assisted crop improvement. Several DNA marker systems like restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), microsatellites or simple sequence repeat (SSR), single-nucleotide polymorphisms (SNP), etc., have been developed. The high-throughput sequencing platforms further enhanced provide thousands of SNPs. They help us to identify genomic region/s or gene/s that control a trait like Zn, Fe, vitamin A, or any nutraceutical. It is followed by reconfirming the gene–trait relationship and their usage options to develop better rice varieties. Transgenic strategies were also applied to develop biofortified grains of rice, especially targeting vitamin A deficiency since 1992 (Yu et al. 2022a). With nutritional quality becoming a primary global breeding target in rice improvement programs across ecologies, marker-assisted selection (MAS) for the significant quantitative trait loci (QTLs) would expedite the development of rice biofortified varieties through marker-assisted breeding (Neeraja et al. 2017). QTL mapping provides opportunities to identify the genomic region (s) associated with the targeted traits by combining genome information with phenotyping. Subsequently, identified genomic region(s)/QTLs/genes could be deployed in the breeding programs through MAS (Collard and Mackill 2008). In marker-assisted backcross breeding (MABB), DNA markers associated with desirable traits are used to select a plant for inclusion in a breeding program. In stage 1, the donor parent having the trait of interest is crossed with the recipient, followed by a screening population with markers to identify the progeny having the trait of interest, which is backcrossed once or repeatedly with the recipient to retain the essential features of the recipient parent (Collard and Mackill 2008). The second stage entails choosing backcross progeny containing the targeted gene using closely spaced flanking markers to reduce linkage drag. The last step is the selection of backcross progeny with background markers (Islam et al. 2017). Hence, a better line with only the donor parent's targeted gene and the recipient parent's remaining genome is obtained (Hasan et al. 2015). In the last few years, either all or the contrasting lines of the trait along with the parents are being subjected to genotype by sequencing (GBS). However, among the several thousands of SNP, a few thousand that are available in both parents are being utilized for the identification of candidate genes that may dilute the purpose of identifying the genomic regions of the donor parent. Genome-wide association mapping is an excellent approach to characterizing the phenotypic and molecular diversity to identify marker–trait associations (Huang et al. 2010). Genomic selection (GS) is another potential strategy

deploying genome-wide markers and phenotypes of phenotypic traits that have been proposed as a recent breeding strategy, especially for the improvement of quantitative traits (Meuwissen et al. 2001). The GS could be useful for developing biofortified varieties in rice as there are major common QTLs/genomic regions associated with nutrients that can be used globally. Among several strategies/approaches toward supporting the development of rice biofortified genotypes, CRISPR-based genome editing appears to be the most promising owing to its targeted gene editing (Kumar et al. 2022). The polymorphic sequence information of genes associated with nutrient-related genes is helpful to check the prudence of the identified SNP through clustered regularly interspaced short palindromic repeats, and CRISPR-associated proteins9 (CRISPR/Cas 9) technology.

2 Nutritional Composition of the Rice Grain

The nutritional composition of rice has been well documented and reviewed, covering both improved varieties and landraces. Rice is the primary source of carbohydrates for regions of the world, consuming it as a staple diet. Mature rice grains are generally harvested with about 20% moisture and dried to 14% moisture for safe storage. The complete grain consists of an outer husk/hull layer (25%), germ and bran layers (10%), and a large endosperm (65%). The resultant product is called brown rice when the husk/hull is removed from the complete, mature grain. Brown rice (100 g) possesses protein (7.3 g), crude fat (2.2 g), carbohydrate (71.1 g), dietary fiber (4 g), and crude ash (1.4 g) with 384 kcal. It has vitamins and mineral content in 100 g in the form of thiamine (0.29 mg), riboflavin (0.04 mg), niacin (4 mg), vitamin E (0.8 mg), Fe (3 mg), and Zn (2 mg) (Souci et al. 1986; Eggum 1969, 1977, 1979). Therefore, brown rice is a good source of energy, fat, mono- and polysaturated fats, carbohydrates, protein, fiber, B complex vitamins, calcium, Fe, and Zn. The B-complex vitamins like thiamin, riboflavin, and niacin promote youthful energy and nourishment to the skin and blood vessels (Lloyd et al. 2000). Vitamin E (especially tocopherols and tocotrienols), selenium, manganese, magnesium, potassium, and manganese are lost during milling and polishing (Lloyd et al. 2000; Lee et al. 2019). This is partly reduced in brown rice with or without polishing. Though brown rice is relatively healthier, polished rice is the preferred form for consumption because of its palatability, cooking ease, and longer shelf life. It is the most preferred form of rice provided through the public distribution system. Populations that subsist on polished rice are vulnerable to vitamin and mineral deficiencies. Various levels of polishing of brown rice give white rice, and bran is a by-product of the brown rice polishing process. However, bran has yet to be fully exploited for its enormous economic and nutraceutical values. Researchers have investigated the anticancer activities of tocopherols (Kline et al. 2004), tocotrienols (Lloyd et al. 2000), and the ability of oryzanols to reduce cholesterol absorption (Lloyd et al. 2000). It was also found that tocotrienol in rice bran can prevent or reverse blood clots and lesions that may lead to strokes or thrombosis (Frei and Becker 2004). In addition, the bran of

red and purple rice is rich in anthocyanins and tannins, which possess antioxidant and anti-inflammatory properties.

Several landraces have been reported to be rich in various nutrient contents in rice such as Zn, Fe, and other minerals, along with bioactive components (Bollinedi et al. 2022; Berni et al. 2018). Recently colored or pigmented rice grains are receiving significant attention because of their higher mineral content, antioxidants, phenolics, and anthocyanins (Raghuvanshi et al. 2017; Shao et al. 2018; Mbanjo et al. 2020). The color of the rice grain is attributed to the accumulation of secondary metabolites like carotenoids and flavonoids (Lap et al. 2021). The health benefits of colored rice grains are well demonstrated through several research studies on rice. Black rice with enhanced anthocyanins is shown to have nutraceutical properties (Butelli et al. 2008; Zhang 2021). The therapeutic properties of colored rice grains, with different local names, in traditional treatments in India and the Philippines have also been widely reported (Ahuja et al. 2007; Umadevi et al. 2012; Hemamalini et al. 2018). However, the existing rice varieties do not provide sufficient micronutrients for optimal health. In this regard, breeding specific nutrients for different market segments is critical. Collaborative DA-PhilRice-IRRI's Heirloom rice project promoted the Cordillera Autonomous Region (CAR) and Arakan Valley Complex (AVC) landraces, especially those that were pink to purple colored rice possessing higher anti-oxidants, and got them premium prices in the market by way of genetic purification, phenotypic characterization, and deeper understanding of their nutritional values (http://cure.irri.org/stories-archive/heirloom-rice-recovering-a-vanishing-treasure). The role of tocopherol and anthocyanin from red and purple-colored rices in cleaving the reactive oxygen species (Ghasemzadeh et al. 2018; Suwannakul et al. 2015) and antioxidation capacity in reducing the inflammatory response (Petroni et al. 2017) has been reported. The enhanced content of γ-oryzanolin-colored rice grains compared to normal rice (Chakuton et al. 2012) and the role of phytosterols in the inhibition of the absorption of cholesterol were also reported (Jesch and Carr 2017). Among the carotenoids in rice, lutein and zeaxanthin are predominant, along with low concentrations of carotenes, lycopene, and β-carotene (Melini and Acquistucci 2017). Red rice grains were reported to consist of relatively higher total tocopherol and tocotrienol than polished red rice grains (Gunaratne et al. 2013; Irakli et al. 2016). Interesting divergence of phenolic acids is seen in different colored rices such as red and black and also in brown and polished grains of rice (Zhang et al. 2015; Shao et al. 2018). Differential content of flavanones was reported in colored and white rice grains (Park et al. 2016; Poulev et al. 2019). Pro-anthocyanidins and catechins are indicated to be predominant in red rice, and anthocyanins in black and purple rice grains (Pereira-Caro et al. 2013; Kim et al. 2014; Zhang et al. 2015). High levels of genetic variation were observed for color differences based on the differential quantity and composition of bioactive compounds (Baek et al. 2015). Differential accumulation of color in grains is regulated by several regulatory proteins and enzymes involved in flavonoid and carotenoid biosynthesis. *MYB* and *bHLH* gene family members are the major regulators of these pathways (Lap et al. 2021).

The starch of rice grain's endosperm is also gaining importance with its critical role in the glycemic index (GI; Butardo and Sreenivasulu 2016). GI is a measure of

glucose release and uptake upon digestion, leading to increased blood glucose content. It is a clinical measure of the tendency of food or drinks containing 50 g of available carbohydrates to influence the blood glucose release and response upon intake relative to the same amount of glucose standard (Jenkins et al. 1981). Rice genotypes show intermediate (>55–69) to high GI (≥70), with a few genotypes with low GI (≤55) (Fitzgerald et al. 2011; Atkinson et al. 2008). Recently, the Indian rice varieties, Improved Samba Mahsuri (Sundaram et al. 2018) and RNR15048 (Prasanthi et al. 2019), have been reported to have a low GI (~51). Polished rice has a higher glycemic load and may impact glucose homeostasis. Genetic manipulation of GI traits was reported to be associated with yield penalty and undesirable cooking and textural properties (Jukanti et al. 2020).

Rice being a staple food, various sections of society – research and development, extension agencies, seed producers, farmers, paddy buyers, rice sellers, and consumers – are associated. Hence, NIN collected samples from 107 selected districts across the country and analyzed a series of compounds in all the samples (Longvah et al. 2017).

The market rice has around 9.3% moisture content. Starch is around 75% and is the dominant component of brown and polished grains, followed by 4.3–20.2% of protein (Chattopadhyay et al. 2011; Santos et al. 2013), 2.4–3.9% of lipids (Juliano 1977), and 0.5–2.0% of other compounds.

2.1 Starch

Starch contains both amylose and amylopectin that assemble to form crystalline large (irregular to cubical), medium or small (spherical or ellipsoidal) granules (Burrell 2003; Reddy and Bhotmange 2013). It comprises 70% amorphous regions of amylose and branching points of amylopectin and 30% crystalline, which contains outer chains of amylopectin (Reddy and Bhotmange 2013). Amylose is a linear or unbranched molecule with α-D-(1-4) linkages between two successive glucose residues. Rice is categorized into waxy (0%), very low (1–9%), low (10–19%), intermediate (20–25%), and high (>25%) amylose (Juliano, 1971). Amylopectin is branched with α-D-(1-4) and α-D-(1-6) glycosidic bonds within the branches and branching points, respectively. The amylopectin proportion decreases with the increase in amylose content.

Sugars are transported from the phloem of the seed coat to the maternal tissue, to the embryonic apoplast, and finally to the endosperm due to the osmotic gradient between the leaf and the growing seed. In the cytoplasm, ADP-glucose pyrophosphorylase (AGPase) activates glucose to ADP-glucose, which enters into the amyloplasts where granule-bound starch synthase I (GBSSI) uses ADP-glucose to synthesize amylose chains. In japonica rice, this enzyme (Wx^b) is less efficient due to a transversion (G to T) mutation in the first intron. Three categories of enzymes coordinate amylopectin synthesis. Soluble starch synthases (SSI, SSIIa, and SSIIIa) elongate (add glucose), and starch branching enzymes (BEI, BEIIa, and BEIIb) create branching points. Debranching enzymes (isomerase and pullulanase)

will enhance the granule compactness by removing some branching points. Involvement of multiple enzymes may accumulate similar amounts of starch; however, they can vary in chain length, branching density, compactness, etc.

Regarding the medicinal/physiological properties and functions of starch concerning human health, in the gastrointestinal tract or gut, starch digestion in the mouth is negligible since the food is generally swallowed into the stomach. Pancreatic juice contains multiple enzymes. The α-amylase and maltase digest starch into maltose and glucose, respectively. SGLT1 and GLUT2 are hexose transporters in Caco-2 cells. GLUT2, a facilitated transporter, is on the apical side and gathers to the membrane at a high glucose concentration. It is the main transporter of glucose (Kamiloglu et al. 2015) into the blood, leading to its increase above normal level (140 mg/100 ml).

In response, insulin hormone is secreted into blood from pancreatic β-cells. Insulin decreases the blood glucose to a normal level by either converting the extra glucose to glycogen in the liver and muscles or into fats and other molecules. During fasting or between the two meals, glucose can decrease by <80 mg/100 ml of blood. In response, glucagon hormone is secreted from pancreatic α-cells and is responsible for glucose release from glycogen and for making glucose from fats and other molecules. Type 1 DM is due to the absence or lower levels of insulin. Insulin is secreted in type 2 DM; however, it cannot show its effect.

Rice varieties were classified into low (GI $\leq$ 55), medium (GI 56–69), and high (GI $\geq$ 70) glycemic index. Lalat and Sampada recorded low GI and gave high, medium, and low glycemic load (GL) for 100, 50, and 25 g raw rice per meal, respectively (Table 1). Therefore, GI and available carbohydrate values must be considered before deciding the approximate amount of low GI rice per meal to maintain safe blood glucose levels following routine day-to-day activities (Sanjeeva Rao et al. 2022).

$$\mathrm{GI} = (\mathrm{iAUCtest\ food}/\mathrm{iAUCreference\ food}) \times 100$$

$$\mathrm{GL} = (\mathrm{GI} \times \text{available carbohydrate in the given amount of food})/100\ (\mathrm{J\ et\ al.\ 1997})$$

The rate of glucose release is more from amylopectin than from amylose, which need more time for digestion, and its consumption may help diabetic patients. Total starch, resistant starch (RS), amylose, and amylopectin contents vary among the rice

Table 1 Available carbohydrate, in vivo GI and GL in rice varieties

S. No	Variety	Available carbohydrate in 100 g rice	GI	GL for 100 g raw rice	GL for 50 g raw rice	GL for 25 g raw rice
1	Sampada	75.28	51.17	38.52	19.26	9.63
2	Lalat	75.84	53.17	40.32	20.16	10.08
4	Savitri	76.52	68.85	52.68	26.34	13.17
5	Jaya	78.53	75.58	59.35	29.68	14.84

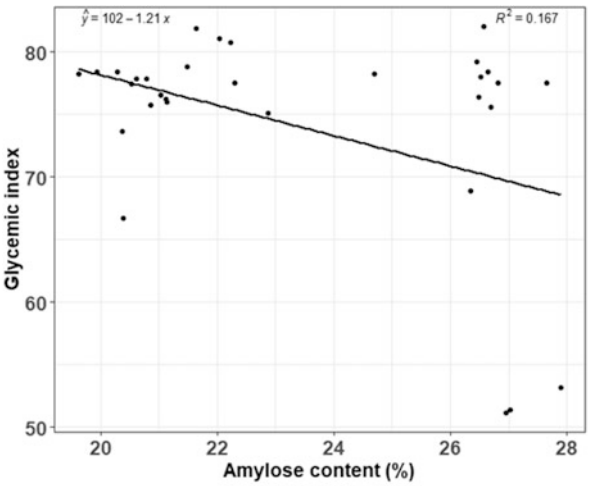


Fig. 1 Relationship between GI and amylose content

varieties and hybrids. Rice varieties released through AICRIP are of intermediate to high amylose. The in vivo glycemic index decreased with the increase in amylose content (Fig. 1). However, varieties with similar amylose content also varied in glycemic index. Amylose content varied from 25.25 to 33.70% for Samba Mahsuri across the locations. Glycemic index can vary with variation in amylose content, particularly varieties with low marginal GI can become medium GI and vice versa. This minor variation may not significantly affect blood glucose.

2.1.1 Resistant Starch (RS)

It was first identified and described as a small fraction of starch that does not hydrolyze to the D-glucose (Englyst and Cummings 1985) within 120 min in the small intestine (Ratnayake and Jackson 2008) and enters the large intestine (Englyst et al. 1992) where it is fermented by microflora (Nugent 2005) leading to flatulence. RS is linear and low molecular weight (1.2 × 105 Da) derived from the retrograded amylose fraction (Tharanathan 2002). In rice, RS range from 0.5 to 2.5%. RS content can vary among humans for the same rice variety due to variations in gastrointestinal functioning, like enzyme concentrations, the viscosity of the gastric fluid, and the food matrix itself. This is the chief reason for the dissimilarity between in vivo and in vitro results.

2.2 Protein

Proteins are distributed both in the aleurone layer and endosperm. They are classified into albumins (16–18 bands with 10–200 kDa), globulins (9 bands with 10–150 kDa), glutelins (12–21 bands with 22–33 kDa), and prolamins (5 bands

Table 2 Protein (%) categories and total protein content in polished grain

Variety	Albumin	Globulin	Prolamin	Glutelin	Total protein
Vibhava	0.54 ± 0.15	0.072 ± 0.03	0.72 ± 0.89	0.54 ± 0.05	4.95 ± 0.37
Ravi	0.38 ± 0.06	0.051 ± 0.006	0.36 ± 0.01	0.45 ± 0.19	6.09 ± 0.74
BPT 5204	0.14 ± 0.07	0.025 ± 0.004	0.81 ± 0.37	0.61 ± 0.40	6.52 ± 0.07
Sampada	0.12 ± 0.03	0.023 ± 0.002	0.28 ± 0.01	0.36 ± 0.17	5.25 ± 0.23
Rasi	0.03 ± 0.003	0.033 ± 0.002	0.82 ± 0.37	0.16 ± 0.09	5.48 ± 0.312

Table 3 In silico amino acid composition of protein sequences

Protein	Number of amino acids	Proportion of EAA %
Albumin	823	51.27
Globulin	155	44.82
Prolamin	237.0	41.5
Glutelin	499.56	43.32

with 10–16 kDa) and their availability (Table 2) varied among the varieties. More number of amino acids were observed in albumins (Table 3), while the proportion of essential amino acid (EAA) was the highest in albumins (Wen and Luthe 1985; Ogawa et al. 1987; Chrastil and Zarins 1992; Umadevi et al. 2019).

In four rices (three inbred varieties – Aghoni Bora, Swarna, and Samba Mashuri, – and one hybrid – DRRH3), among the nonessential amino acids (NEAA), glutamic acid was the highest and cysteine was the least. Of the EAA, leucine and arginine were the highest while the least available amino acids varied among the samples (Fig. 2).

Almost all the rice protein is digestible and less immunogenic. Hence, it is fed to infants. Protein digestion starts in the stomach where pepsin breaks proteins into peptides. In the small intestine, chymotrypsin and trypsin further hydrolyze peptides to oligopeptides that are hydrolyzed into amino acids by exopeptidases. Proteins as such cannot be absorbed in the gut and are the only sources of amino acids (protein synthesis in humans) that are absorbed into the blood.

Ten of the 20 standard amino acids are EAA since they cannot be synthesized in the body (Table 4) and must be supplied through food. Glutamic and aspartic acids provide an amino group for the synthesis of NEAA. However, histidine and arginine are essential for growing children and pregnant women. Cysteine and tyrosine are synthesized from methionine and phenylalanine, respectively.

Protein turnover is the amount of protein synthesized and degraded per day. It is around 2% in humans and can increase during pregnancy. About 75% of the amino acids released during degradation are reutilized, while the remaining are lost as urea. The intake and output of protein is equal in nitrogen equilibrium. The intake is less while output is higher and vice versa in the negative (leads to kwashiorkor, especially in children) and positive nitrogen balance, respectively.

In India, so far, three high-protein (>10%) varieties – CR Dhan 310, CR Dhan 411, and CR Dhan 311 in milled rice – were released through the AICRIP biofortification trial and one BRRI Dhan 84 (9.7%) was released in Bangladesh.

The average daily requirement of nitrogen is 0.83 g/kg body weight, and the average body weight is 65 kg for India, thus, 51.46–53.95 g protein (72 g for pregnant or breastfeeding women) is required. These high-protein varieties can

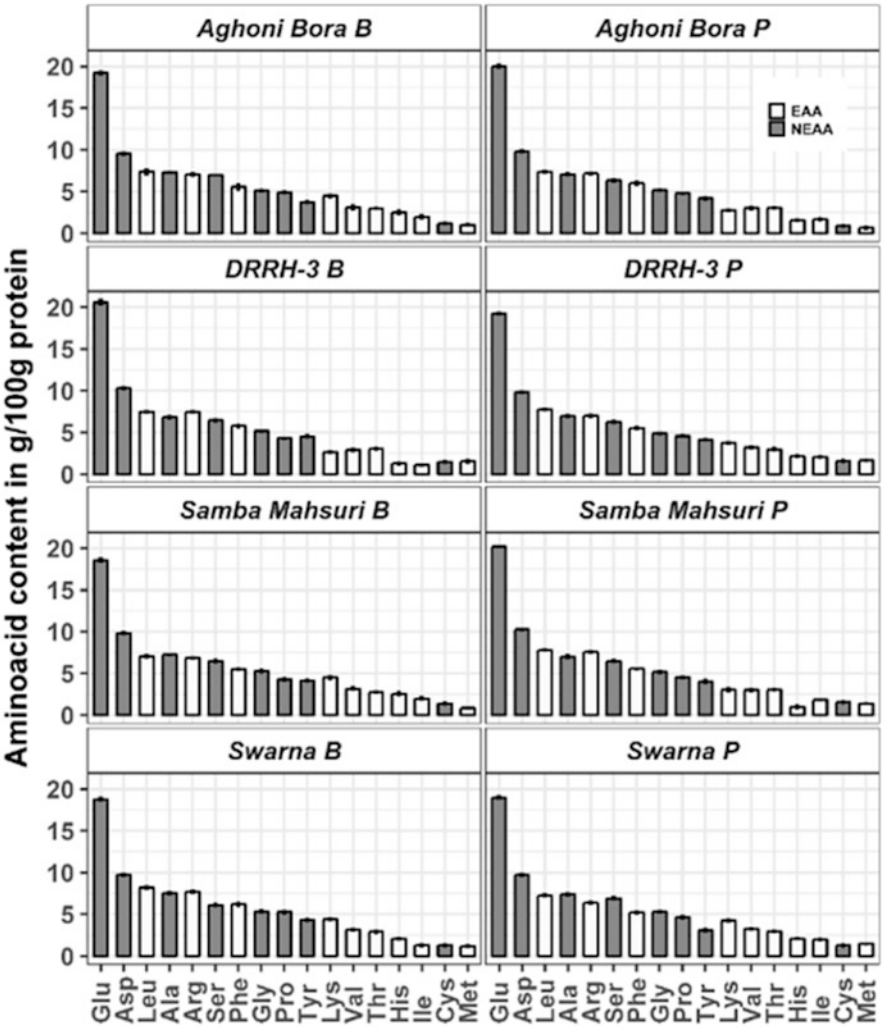


Fig. 2 Amino acids (17) in brown (B) and polished (P) rice (unpublished data)

provide around 24–30 g of protein from 220 g of rice. As nitrogen fertilizer application decreases in the future, germplasm screening must be intensified to identify donors and promising markers to enhance protein.

2.3 Lipids

Lipids are present in the aleurone layer (marginally higher) and endosperm. Bran is a by-product while milling brown rice. Lipids extracted from bran are called bran oil, containing triglycerides, free fatty (bulky side chains) acids, and oryzanol. Rice contains

Table 4 EAA (g/100 g) in raw rice (Longvah et al. 2017) and their role in human

Amino acid	Brown	Milled	Role
Methionine (M)	2.39 ± 0.26	2.60 ± 0.34	Increases the antioxidant levels (glutathione). Reduce blood cholesterol level
Arginine (R)	7.69 ± 0.37	7.72 ± 0.55	Precursor for nitric oxide, ornithine, polyamines, agmatine, proline, glutamate, creatine, and urea. Optimal growth and development of infants
Threonine (T)	3.38 ± 0.25	3.28 ± 0.27	Prevents fatty buildup in liver
Tryptophan (W)	1.00 ± 0.17	1.27 ± 0.14	Prevents fatty buildup in liver. Precursor of neurotransmitter serotonin (calming effect)
Valine (V)	6.72 ± 0.36	6.06 ± 0.02	Influences brain to uptake of tryptophan, phenylalanine, and tyrosine
Isoleucine (I)	4.08 ± 0.51	4.29 ± 0.23	Formation of hemoglobin. Prevents muscle wasting in debilitated individuals
Leucine (L)	8.40 ± 0.55	8.09 ± 0.40	Promotes healing of skin and broken bones. Reduces protein breakdown in the muscles
Phenylalanine (F)	5.50 ± 0.49	5.36 ± 0.43	Production of collagen. Precursor of tyrosine. Enhances learning, memory, mood, and alertness
Histidine (H)	2.36 ± 0.18	2.45 ± 0.30	Production of RBC (anti-anemic) and WBC
Lysine (K)	3.63 ± 0.29	3.70 ± 0.39	Inhibits viruses – herpes simplex virus. Lysine and vitamin C together form L-carnitine that enables muscle tissue to use oxygen more efficiently and delay muscle fatigue

higher levels of C16 and C18 fatty acids. Notably, the latter was at the highest level of the two essential fatty acids, linolenic and linoleic acids. Eventually, the proportion of unsaturated or polyunsaturated fatty acids (PUFA) is higher (Table 5).

The growing fatty acid gains 2 carbons (acetyl CoA) during each condensation cycle, and it grows up to C16 in the cytoplasm. Further elongation as well as unsaturation (introduction of double bonds) occurs in the endoplasmic reticulum.

Oryzanol was first isolated by Kaneko and Tsuchiya (1955). It is a mixture of sterol esters of ferulic acid. It contains cycloartenol, β-sitosterol, 24-methylene-cycloartanol, cyclobranol (cycloartenol), and campesterol (4-desmethysterols) (Rogers et al. 1993), and composition varies among varieties.

Regarding the medicinal/physiological properties and functions of lipids of rice concerning human health, triacylglycerols are emulsified by bile juice into micelles. Pancreatic lipase digests them into free fatty acids and glycerol, which enter into intestinal cells where they are resynthesized into triacylglycerols and packaged with cholesterol and specific proteins to form chylomicrons.

Chylomicrons enter into the lymph and, through blood, reach various tissues to supply fatty acids and glycerol. The remnant chylomicrons reach the liver and are converted into very low-density (VLDL), low-density (LDL), or high-density (HDL) lipoproteins based on the availability of lipids.

Table 5 Fatty acids in raw rice samples collected across the country (Longvah et al. 2017)

S. No	Compound (mg/100 g)	Systemic name	Brown	Milled
1	Myristic	Tetradecanoic acid	30.42 ± 3.15	13.19 ± 3.00
2	Palmitic	Hexadecenoic acid	273 ± 14.9	143 ± 28.0
3	Stearic	Octadecanoic acid	33.01 ± 4.34	14.50 ± 3.27
4	Arachidic	Eicosanoic acid	3.09 ± 0.21	1.46 ± 0.40
5	Behenic	Doeicosanoic acid	1.98 ± 0.21	1.98 ± 1.49
6	Lignoceric	Tetraeicosanoic acid	2.49 ± 0.38	1.14 ± 0.35
7	Palmitoleic	*cis*-9-Hexadecenoic	2.77 ± 0.46	1.49 ± 0.47
8	Oleic	cis-9-Octadecenoic	197 ± 15.4	109 ± 21.2
9	Eicosenoic	icos-11-enoic acid	1.89 ± 0.25	1.54 ± 0.44
10	Nervonic	Tetracos-15-enoic acid	0.89 ± 0.22	0.75 ± 0.36
11	Linoleic	all-cis-9,12-Octadecadienoic	490 ± 33.2	234 ± 45.8
12	α-Linolenic	all-cis-9,12-Octadecadienoic	16.10 ± 0.92	9.51 ± 1.09
13	Total saturated fatty acids (SFA)		346 ± 20.3	184 ± 8.9
14	Total mono-unsaturated fatty acids		203 ± 15.7	117 ± 6.6
15	Total PUFA		506 ± 33.6	253 ± 13.2

In the liver, PUFA is converted into ketone bodies. While SFA is converted into LDL (Beynen and Katan 1985) and acetyl CoA, which is converted into cholesterol, this quantity is different from the one absorbed from the intestine where γ-oryzanol in food can inhibit cholesterol absorption. Rice contains more linolenic acid and will reduce LDL and simultaneously increase the HDL (Cicero and Gaddi 2001).

Linolenic acid, the precursor of eicosanoids (C20) and docosahexaenoic acid (DHA; ω3, 22:6), is converted into eicosanoic acid (C20) that in turn form prostaglandins and thromboxanes (hormones) and into leukotrienes and lipoxins. Prostaglandins deal with inflammation and regulate blood coagulation and reproduction. Leukotrienes help in muscle contraction and chemotactic properties and are slow reactive substances of anaphylaxis (SRSA). DHA is required for brain and retina development. Thus, except for essential fatty acids, human metabolism is capable of synthesizing all other lipids from protein or carbohydrates.

2.4 Other Compounds

Rice also contains vitamins (Table 6), minerals, phytosterols, organic acids, etc. Some are more in the aleurone layer (Fe), and others are more in the endosperm. Highly pigmented brown rice is in red, purple, or black color (Table 7), and the various concentrations of cyanidin-3-O-glucoside equivalent (CGE) and catechin acid equivalent (CAE) are responsible for the variation in the color (Goufo and Trindade 2014). Generally, pigments (flavonoids) are glycosylated (aglycone), methylated or acylated forms of anthocyanidin ("Anthos" means flower, and "kyanos" is blue). Pigmented rice contains carotenoids, 159 and 16.87 μg/100 g in brown and milled rice, respectively.

Table 6 Water-soluble vitamins in raw rice (Longvah et al. 2017) and their role in human

S. No	Compound (mg/100 g)			Functions	Deficiency symptom
		Brown	Milled		
1	Thiamine (B1)	0.27 ± 0.023	0.05 ± 0.019	Coenzyme in pyruvate, α-ketoglutarate, dehydrogenases, and transketolase; nerve conduction	Beri-beri central nervous system lesions
2	Riboflavin (B2)	0.06 ± 0.011	0.05 ± 0.006	Coenzyme in redox reactions. Prosthetic group of flavoproteins	Lesions of corner of mouth, lips, tongue; seborrheic dermatitis
3	Niacin (B3)	3.40 ± 0.12	1.69 ± 0.13	Coenzyme in redox reactions. Formation of NAD and NADP	Pellagra
4	Pantothenic acid (B5)	0.61 ± 0.04	0.57 ± 0.05	Functional part of CoA and ACP in fatty acid metabolism	Huntington's disease (HD)
5	Total pyridoxine (B6)	0.37 ± 0.035	0.12 ± 0.012	Coenzyme in transamination, decarboxylation, glycogen phosphorylase, and steroid hormone action	Disorders of amino acid metabolism, convulsions
6	Biotin (B7) (μg)	1.38 ± 0.21	0.60 ± 0.12	Coenzyme in carboxylation, gluconeogenesis, and fatty acid synthesis	Impaired fat and carbohydrate metabolism, dermatitis
7	Total folates (B9) (μg)	11.51 ± 1.69	9.32 ± 1.93	Coenzyme in one carbon transfer	Megaloblastic anemia

Table 7 Composition of colored rice

Pericarp color	Anthocyanin (mg/100 g)	Proanthocyanidin (mg/100 g)
Black	1884	78
Purple	2874	525.4
Red	8.78	716.6
Brown	3.09	4.34

Phenylalanine is converted into 4-coumaroyl CoA to chalcone (committed step) in the presence of ATP, CoA, and three molecules of malonyl CoA (Fig. 3). Chalcone is converted into naringenin, which is oxidized to dihydrokaempferol. It, in turn, is hydroxylated at position 3 or 5 to dihydroflavonols, which are reduced to

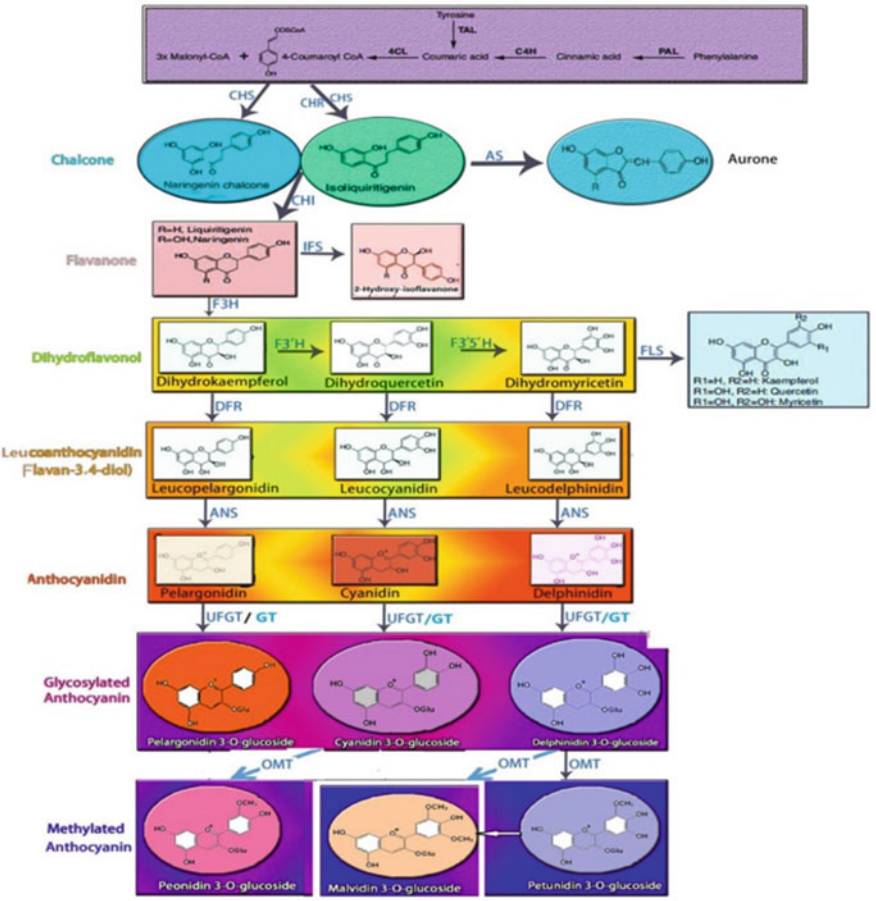


Fig. 3 Anthocyanin biosynthetic pathway (Mackon et al. 2021)

leucoanthocyanidin by NADPH and dihydroflavonol 4-reductase (DFR), which is another critical enzyme. The leucoanthocyanidin is oxidized to anthocyanidin and sequentially glycosylated and methylated to form anthocyanins (Mackon et al. 2021).

The chalcones are isomerized to flavanones and converted into flavones by flavone synthases (FS I and II). FSII is a group of cytochrome P450 enzymes (CYP93G1) that desaturate naringenin to apigenin (Fig. 4), which is hydroxylated to luteolin followed by esterification to form chrysoeriaol, hydroxylation to form selgin, and final esterification to form tricin (Lam et al. 2015).

Regarding the medicinal/physiological properties and functions of other compounds of rice concerning human health, the Ayurvedic treatise indicates the prevalence of medicinal rice varieties in India (Das and Qudhia 2001). Both Njavara and Jyothi are red rice; however, only Njavara is considered to have medicinal properties (circulatory, respiratory, and digestive systems). Antioxidant contents and capacities of pigmented rice were much higher than white rice (Pramai and Jiamyangyuen

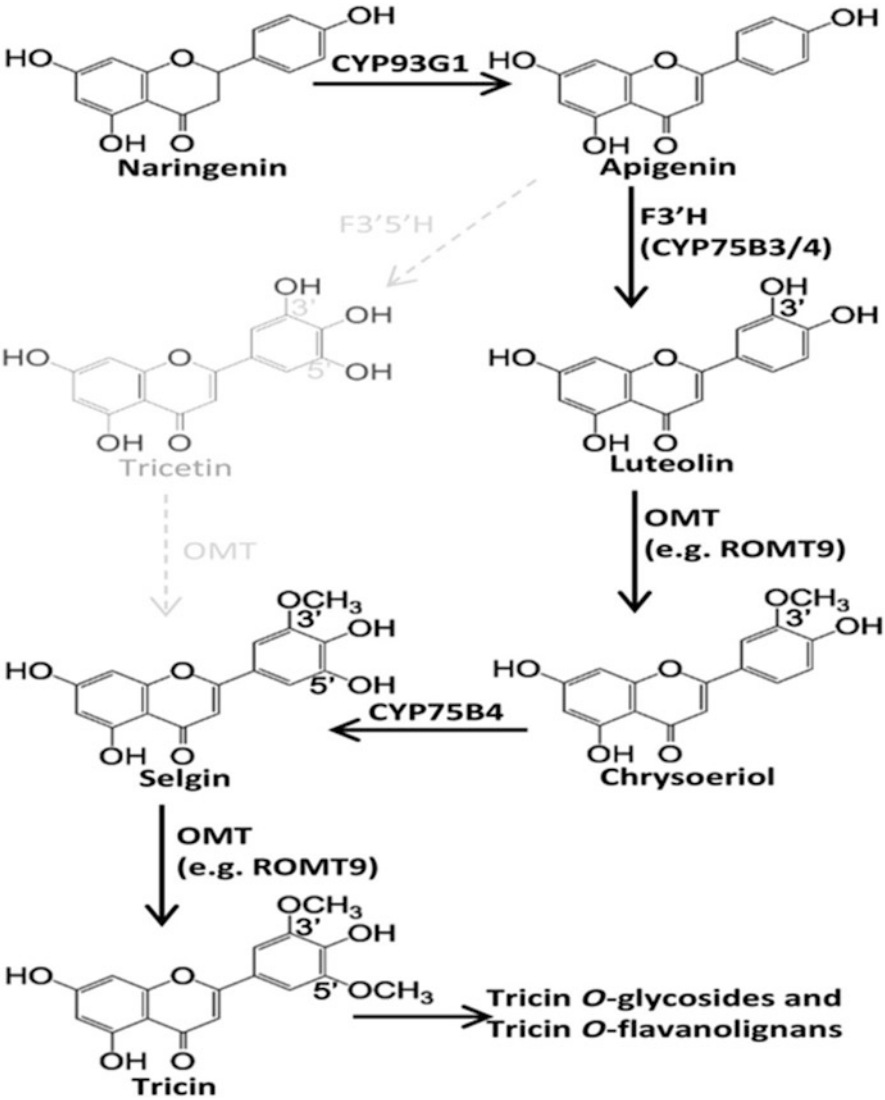


Fig. 4 Tricin biosynthesis pathway (Lam et al. 2015)

2016). Njavarakizhi is prepared by cooking Njavara rice in milk, with herbs, like *Sidarectusa* and *Alpinia galanga,* for the treatment of paralysis, arthritis, and neurological problems (Das and Qudhia 2001). Antioxidant, antiarthritic, antidiabetic, and antigastritis (peptic ulcers) activities were observed in Kavuni rice (Valarmathi et al. 2015). Karungkavuni is useful to cure elephantiasis and contains antioxidant, hypercholesteremic, hepatoprotective, anti-inflammatory, cancer-preventive, and antimicrobial compounds (Kalaivani et al. 2018).

In black rice (Longjin), cyanidin-3,5-diglucoside is the major compound (Hou et al. 2013). Ethanol extract of black rice bran (EEBRB) contains 3.28 ± 0.34 mg/

100 g of anthocyanin content, which can regenerate pancreatic β-cells (Wahyuni et al. 2016), particularly cyanidin-3-glucoside (Tantipaiboonwong et al. 2017) which competes with glucose (having glucose moiety) to get absorbed through GLUT2 (Kroon et al. 2004) and inhibits glucose absorption (Kamiloglu et al. 2015).

Despite the effector compound yet to be identified, extracts of black or red rice can inhibit the multiplication of breast cancer cells (Ghasemzadeh et al. 2018) by inhibiting cytochrome P450 or scavenging free radicals (Insuan et al. 2017). Similarly, purple rice extract was effective against hepatocarcinogenesis (Suwannakul et al. 2015). Tricin has been noted to have antioxidant, anticancer, anti-inflammatory, and cardiovascular properties (Lam et al. 2015). It inhibits cyclooxygenase I, which leads to the reduction of prostaglandin E_2 (PGE_2) and the inhibition of colorectal cancer cells.

The vitamin E tocotrienols (0.08 mg/100 g) and tocopherols (1.09 mg/100 g) in nonpigmented rice (Longvah et al. 2017) are lesser than in pigmented rice (Irakli et al. 2016). The antioxidant activity of oryzanol is due to ferulic acid moiety. It inhibits cholesterol oxidation and lipid peroxidation in retinal homogenates under oxidative stress (Hiramitsu and Armstrong 1991) and can kill leukemia cells (Parrado et al. 2006). The scavenger receptor class B type I (SR-BI) on enterocytes, CD36 membrane protein at the border of duodenum and jejunum, and NPC1-like transporter 1 (NPC1L1) is a major sterol transporter in the intestine that help in the uptake of carotenoids, fat-soluble vitamins, long chain fatty acids, etc. (Reboul 2019).

The ecological and climatic conditions were reported to influence the anthocyanin and bran oil content. Anthocyanin content was higher in two genotypes in the lowland and others in the highland. Similarly, antioxidant capacity was fourfold higher in lowland and other in highland (Rerkasem et al. 2015). Rice experiences high temperature during the grain filling stage of the dry season. In the AICRIP trial, high-temperature conditions (4–5 °C higher than the control) were created, covering the treatment with a polythene sheet. In eight entries, bran oil content ranged from 10.5–12.5 and 5.8–13.0 g/100 g bran in control and high-temperature stress, respectively. The octanoic acid peak was higher while the palmitoleic acid peak was smaller under treatment in susceptible.

2.5 Iron (Fe) and Zinc (Zn)

Around 3.1% and 0.9% of 3177 germplasm showed Zn content ≥35 and ≥40 mg/kg, respectively (Sanjeeva Rao et al. 2020). As the maximum Zn content in germplasm is four- to fivefold higher than the cultivars, high Zn varieties (24 mg/kg) were released (Table 8).

For Fe, diversity in germplasm itself is narrow, and moreover, 70% is lost during milling. Hence, biotechnological tools were used to release 1 (the Philippines), 1 (Latin America), and 3 (Bangladesh) varieties having high Fe of 7–10 mg/kg in polished and 13–31 in brown rice.

Table 8 List of high Zn (mg/kg) biofortified rice varieties released

S. No	Country	Name of varieties	Zn	Sl. no.	Country	Name of varieties	Zn
1	India	DRR Dhan 45	22.3	13	Bangladesh	Binadhan 20	26.5
2		Chhattisgarh Zn Rice 1	22–24	14		BRRI Dhan 62	19.6
3		DRR Dhan 48	20.91	15		BRRI Dhan 64	25
4		DRR Dhan 49	26.13	16		BRRI Dhan 74	24.2
5		Surabhi	22.84	17		BRRI Dhan 84	27.6
6		GR-15	21.58	18		BU Aromatic Hybrid Dhan-1	21.8
7		Surabhi	22.84	19		BU Aromatic Dhan-2	22
8		Zno Rice-MS	27.4	20	Indonesia	INPARI 47 Nutri Zn	29.54
9		Chhattisgarh Zn Rice-2	22–24	22		INPARI IR Nutri Zn	26
10		DRR Dhan 63	24	23	LAC	CENTA A-Nutremas	22.86
11		CR Dhan 315	24.9	24		CIAT BIO-44 +Zn	25
12	Philippines	NSICRc460	19.6			Fedearroz BIO Zn 035	26

The recommended dietary allowance (RDA) of Fe and Zn for the human population (25–50 years of age) are 10–15 and 12–15 mg, respectively. In India, the average daily intake of rice is 220 g, and polished rice having 45.5–68.2 mg/kg Fe and 54.5–68.2 mg/kg Zn can only meet the RDA without considering the bioavailability (Sanjeeva Rao et al. 2014).

As the maximum Zn in brown rice is 45 mg/kg, it may not be possible to get varieties >50 mg/kg through conventional breeding coupled with Zn-deficient soils. Strategies are to be chalked out to test rice's actual genetic potential with sufficient Zn experimental fields. If the grain Zn content is still <50 mg/kg, then biotechnological intervention is required.

The highest amount of Zn (50 mg/kg) and Fe (14.7 mg/kg) contents were achieved by NASFer-274 transgenic event having rice nicotianamine synthase (OsNAS2) and soybean ferritin (SferH-1) genes into IR64 controlled by glutelin promoter (Trijatmiko et al. 2016). On crossing this with three popular varieties, the Zn and Fe contents in F2 bulked milled rice ranged from 33–45 mg/kg and 6–11 mg/kg, respectively. As IR64 is also a popular variety in India, this transgenic event may be utilized through MABB to further enhance Fe and Zn contents in the biofortified high Zn varieties.

Soils (micronutrient source) are 30% and 20% deficient in Fe and Zn, respectively. In the AICRIP Biofortification trial, Zn content in high Zn variety Fe and Zn varied from 2.65–7.05 mg/kg and 13–27.85 mg/kg across 17 locations, respectively. Of the seven rice varieties cultivated in low N, N50, and N100, minor variations were noted for grain Fe content except Vandhana (Fig. 5). But grain Zn content was

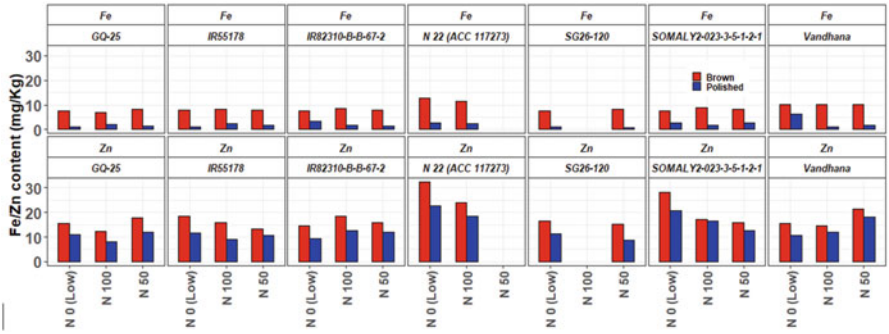


Fig. 5 Fe and Zn content in rice grain at varied soil N

more in Somaly2-023-3-5-1-2-1 and N22 under low N. Purple rice noted higher Zn (33–35 mg/kg) in lowland over highland, and the difference between altitudes ranged from 16 to 50% among the ten genotypes.

Postharvest processing, like parboiling, increases head rice recovery during milling and nutrient content in the endosperm. As most nutraceuticals are heat or acid-sensitive, enzymatic treatment is the best (Parrado et al. 2006), where bran is treated with an endoprotease mixture that hydrolyzes lipase completely to stop rancidity (unpleasant odor) and enhances oryzanol (3.4 times). Alternately, brown rice is packaged in smaller quantities in the presence of nitrogen. The absence of oxygen can avoid rancidity and enhance the shelf-life. However, this will hold well till the packet is opened. Industrial chemical fortification of broken rice powders with Fe and Zn is followed by conversion into rice pellets through an extruder, mixed with field-grown rice, and released into markets.

3 Marker-Assisted Breeding for Health-Related Traits

For achieving food and nutritional security, increasing the production of nutrient-rich rice is challenging. As conventional breeding is time consuming and very dependent on environmental conditions, marker-assisted breeding (MAS) is one of potential modern molecular tools for nutrient improvement of crops. It aids in value addition of varieties by incorporating gene responsible for enhanced micronutrient. By availing technologies like DNA markers, genetic engineering, etc., it helps us to detect the allelic variation in genes underlying the traits. Information on nutritional traits such as grain protein, vitamins, minerals, glycemic index value, and phenolic and flavonoid compounds has been reported (Mahender et al. 2016). Several quantitative trait loci (QTLs) for nutrient traits, amino acids, fat content, low phytate content, etc., have been mapped in rice in different mapping populations (Thangadurai et al. 2020; Raza et al. 2019). Research studies involving wild species, landraces, and mutants in backcrossing breeding programs by applying MAS resulted in the identification of promising lines with high Zn, Fe, high protein, and low phytate content.

3.1 Germplasm Characterization

The importance of genetic resources is widely recognized. There is rich variability in rice genetic resources related to nutritional and health-related traits. Due to a lack of availability of information and limited production, these resources could only reach some of the population. The information on all genetic resources of rice enriched with nutritional and health-promoting traits could be used to develop functional food and drugs to prevent and treat several human disorders, metabolic diseases, oxidative stress-related pathogenesis, and carcinogenesis. Many nutraceuticals are available in rice, and huge germplasm is yet to be screened, which may contain a higher concentration of the known or novel nutraceuticals. Hence, the requirement of biotechnology intervention can be nutraceutical dependent. In other words, a significant amount of research was conducted on Fe, Zn, and golden rice to enhance the same in the rice grain; however, the progress is different.

3.1.1 Rice for Grain Protein and Amino Acids

Many landraces with high crude protein are ARC10063 (16.41%), Kullakar (12.88%), Kempunellu (12.58%), Ambemohar-1 (12.50%), Ambmohar-2 (12.42%), Antarsali (12.31%), Balam rice (B2: 12.4%) of Bangladesh, black rice genotypes of Southern India, viz., BPT 2848 (13.11%), and BPT 2858 (12.82%) (Hanamaratti et al. 2015; Ahmed et al. 2015; Chattopadhyay et al. 2011; Velprabakaran et al. 2020; Sridevi et al. 2021). Two high-yielding varieties with high grain protein content are CR Dhan 310 (10.3% in milled rice), CR Dhan 311 (Mukul) with protein 10.1% in milled rice, and moderately high Zn (20 ppm) (Chattopadhyay et al. 2019).

Some mutant lines with enhanced amino acids are Mtr1 (mtr1-D) mutant rice of Norin-8, with increased accumulation of tryptophan (11.4–16.1%), and mutant of OASA2 (subunit of anthranilate synthase) in Nipponbare with 230-fold higher level of tryptophan (Yamada et al. 2008; Saika et al. 2011). A weedy rice morphotype of *Oryza sativa* from North Macedonia weedy accession, WR1323232, reported having high protein (12.12%) (Dimitrovski et al. 2021). In wild rice species, *O. rufipogon* of the primary gene pool, *O. officinalis* of the secondary gene pool, and *O. meyeriana* of the tertiary gene pool having a high crude protein content of 14.87 mg/100 mg, 15.12 mg/100 mg, and 15.56 mg/100 mg dry seed weight, respectively, were reported (Jiang et al. 2014).

3.1.2 Rice for Vitamins

Particular black rice landraces like Huggi Bhatta, Kalo Bhat, Tiki, and Chak Hao Poireithon were reported to contain the highest concentration of total carotenoid, the precursor of vitamin A (32.86 μg/g; 29.73 μg/g; 21.19 μg/g; 15.39 μg/g,respectively), implying their therapeutic potential in human health and nutrition (Roy et al. 2021). Golden rice, a genetically engineered rice introgressed with two major genes of β–carotene biosynthetic pathway (phytoene synthase from maize, β-carotene desaturase from bacteria *Erwinia uredovora*) into rice endosperm, is another source

of vitamin A. Golden rice 2 contains as much as 37 μg total carotenoids per gram of dry grain weight, of which 31 μg/g is β-carotene (Paine et al. 2005).

Rice landraces with higher vitamin E are some of the Japanese varieties such as Hirayama, Moritawase, and Kaneko that were found to have the highest vitamin E or tocotrienol (80–86%). Landraces of Northeast Thailand, such as Jao Dang (δ-tocotrienol: 8.87 mg/kg), Khao Hawm (γ-tocopherol: 9.78 mg/kg), Tan gun (α-tocopherol: 6.29 mg/kg), Nahng Boon Ma (γ-tocotrienol: 21.6 mg/kg), Ma phai (α-tocotrienol: 9.94 mg/kg), Indian rice landrace like Mapillai samba (26.73 ± 0.49 μg/g), were reported to have the highest vitamin E content (Sookwong et al. 2007; Sudtasarn et al. 2019; Rajendran and Chandran 2020). Mutant varieties with enhanced levels of tocopherols are Shangshida No. 5′ with 2.2-fold and MRXII mutant rice lines of japonica rice cultivar 'Dongan' with 23% higher tocopherol level than that in wild type (Wang et al. 2013; Hwang et al. 2014).

Rice landraces with maximum content of different water-soluble vitamins are Kalajirajaha (0.38 mg thiamin/100 g), Hati bandha (0.09 mg riboflavin/100 g), Kutchi gism (3.87 mg niacin/100 g), Mimittimmidokru (3.1 mg pantothenic acid/100 g), and Mimittimsokmil/mim-kudep (0.17 mg pyridoxine/100 g) (Longvah and Prasad, 2020). Further, the medicinal rice landrace Kullakar with high thiamine content (0.53 ± 0.01 mg/100 g) and Karikalaveya with high riboflavin and niacin content (0.08 ± 0.01 mg/100 g and 16 ± 0.58 mg/100 g, respectively) were identified as a rich source of vitamins (Isaac et al. 2012).

3.1.3 Rice for Minerals

Some of the essential landraces and genotypes with high Fe and Zn and other minerals in brown rice are Maubinni (Fe: 27.3 ppm), Madhukar (Fe: 17.3 ppm; Zn: 53.7 ppm), Thanu (Zn: 55 mg kg^{-1}), and NDR-6279 (Fe: 45.1 mg kg^{-1}); Fe-dense (≥40 ppm) genotypes are Tikimahsuri (52.15 ppm), Jabaphulla (52.15 ppm), Kala Kusuma (52.1 ppm), CR 2327-23 (51.4 ppm), Budhidhan (51.15 ppm), Kalamakhi (50.15 ppm), Nikipankhia (47.2 ppm), ORM 405-8 (45.05 ppm), Jadumani (42.75 ppm), Basudha (41.45 ppm), Malliphulajhuli (41.35 ppm), and Tulasibasa (40.35 ppm); Kuttadan, a landrace of Kerala, was reported being rich in Zn (30.63 ppm) (Longvah et al. 2012; Anuradha et al. 2012; Jagadeesh et al. 2013; Pillai et al. 2020; Tripathy and Dash 2021). The black rice landrace, 'Chak Hao', also harbors many elements (Al, Si, Mg, Cl, K, Ca, P, and S) (Tripathy and Patra 2021).

In polished rice, landraces like Chittimutyalu (Fe: 14.0; Zn:25.7 ppm), Munga (Fe: 15.8; Zn: 28.7 ppm), Basmati 386 (Fe: 13.1; Zn: 27.7 ppm), Ranbir Basmati (Fe: 10.4; Zn: 30.9 ppm), deep water rice Madhukar (Fe: 11.2; Zn:24.2 ppm), TKM 9 (Zn: 24.7 ppm), Thanu (Zn: 23.50 ppm), IR36 (Fe: 23.0 ppm), MTU 3626 (Fe: 19.72 ppm), and NDR 6279 (Fe: 19.60 ppm) were reported as elite Fe, Zn rice genotypes (Ravindra Babu 2013; Jagadeesh et al. 2013).

High-yielding high Zn rice varieties of Bangladesh are BRRI Dhan 62 (19 ppm), BRRI Dhan 64 (25 ppm), high Fe variety, and BRRI Dhan 48 (14.03 ppm) (Khatoon and Islam 2020). High Zn high-yielding varieties of India are DRR Dhan

45 (22.6 ppm), DRR Dhan 48 (24.0 ppm), DRR Dhan 49 (25.2 ppm), Zinco rice MS (27.4 ppm), CR Dhan 311 (protein: 10.1%; Zn: 20.1 ppm), and CR Dhan 315 (Zn: 24.9 ppm) (Yadava et al. 2020). Mutant lines with a rich source of Zn in brown are P44 mutant Sel.-1 and OR CZ 75-3 (51.95 ppm) (Tripathy and Dash 2021).

Wild species of rice such as *O. nivara*, *O. rufipogon* of the primary gene pool, *O. latifolia*, *O. officinalis* of the secondary gene pool, and *O. granulate* of the tertiary gene pool contain around two to three-fold higher Zn than in the cultivated rice, with Zn concentration varying from 37 to 55 mg/kg in nonpolished grains (Anuradha et al. 2012).

3.1.4 Rice for Low Phytic Acid

Some of the low phytic acid (lpa) landraces with low seed phosphorous levels are P 1490-03(3.48 mg/g), *indica* rice cultivar ZN 60 (3.99 mg/g), Yamukh (2.03 mg/g), and Amkei (2.13 mg/g) landraces of Arunachal Pradesh (Wang et al. 2011; Longvah and Prasad 2020; Gyani et al. 2021).

Improved rice varieties such as Goldhull, with a 45% reduction in phytic acid (PA); OM4498, OM5731, DS2000 (mean: 0.465 μg P); Basmati rice, with 54–63% reduction in phytic acid are also reported as low PI rice (Rutger et al. 2004; Lang et al. 2007; Qamar et al. 2019). Low lpa mutant lines are 'Kaybonnet' lpa1–1, with 45% reduction in bran PA with high Zn, mutant lines of MIk gene with 34–64% reduction in PA, Os-lpa-XS110-1 developed from wild type variety 'Xiushui 110' with 46% reduction in PA, N15-186 and N15-375 line with 75% and 43% reduction in seed PA, Pusa LPA Mutant 11 (PLM11), with 70% reduction in PA, Os-lpa-XS110-1 (japonica variety: Xiushui110) with 45.30 mg/kg of Zn and 4.06 mg/g PA compared to wild type (Liu et al. 2004, 2007; Frank et al. 2007, 2009; Kim et al. 2008; Gyani et al. 2021; Wang et al. 2021).

3.1.5 Rice with High-Resistant Starch

Many rice landraces with high-resistant starch (RS) are Kataribhog, Sadanunia, and Chakhao, with RS of 45.72%, 60.49%, and 59–61%, respectively (Mondal et al. 2021). Many mutant rice lines with highly resistant starch areami-BEIIb with an amylose content of 41.2% were obtained through mutation in starch branching enzyme IIb; sbe3-rs mutant line with 60.4% of RS obtained by crossing japonica mutant 'Jiangtangdao 1' with an indica cultivar 'Miyang 23'; ss3a/sbe2b double mutant-resistant starch lines with high amylose content (45%); IR 36 is another line with resistant starch of 44%; be1 be2b is double mutant rice cultivar with an amylose content of 51.7% (Butardo et al. 2011; Yang et al. 2012; Asai et al. 2014; Guzman et al. 2017; Miura et al. 2021).

3.1.6 Low Glycemic Index Rice

Low glycemic index (GI) landraces are CGMD-55 belonging to the Gurmatiya group with a GI value of 55. Karuthakar Poha and Chak Hao Poreiton have a low GI of 52 and 53, respectively (Chandel et al. 2016; Rajendran and Chandran 2020). Improved cultivar such as Laila, a Basmati rice, Australian Doongara rice,

Sinandomeng of Philippines, high-yielding varieties of Bangladesh, BRRI16, BRRI dhan46, and BRRI dhan69; Bangladesh black rice, BBK1 (GI values 48 and 51 in brown and mulled rice); Sona Masoori (GI:51), improved Samba Mahsuri (GI: 50.99) and Madhuraj (GI: 55); two Thailand varieties, PK + 4#20A09 (GI:48.1) and PK + 4#1_E06 (GI:54.6); Indonesian rice cultivar, Logawa GI (54.39 ± 1.49); red rice of South India Karungkuruvai, Kavuni, Kalanamak, and Kullakar are low GI (50–55); GQ02497 with GI value 51.1; the line MAKRO:IRGC 74,763-1 with low GI of 52.91; redfragrant, a variety of Sri Lanka with GI value of 47; Lalat (GI: 53.17), BPT 5204(GI: 51.42), and Sampada (GI: 51) are with low GI values (<55) (Sun et al. 2018; Fitzgerald et al. 2011; Tripathy et al. 2017; Shozib et al. 2017; Shozib 2018; Sarkar et al. 2018; Sundaram et al. 2018; Nounmusig et al. 2018; Kusmiyati et al. 2019; Pushpam et al. 2019; Anacleto et al. 2019; Selvaraj et al. 2021; Abeysiriwardena and Gunasekara 2020; Sanjeeva Rao et al. 2022). Mutant rice line, Frontiere, a new breed of low GI (41) and high protein rice (11%), sold as Parish Rice and Cahokia Rice, has the lowest glycemic index of any rice (Wenefrida et al. 2017).

3.1.7 Phenolics and Flavonoids

Superior landraces enriched with the phenolics and flavonoids are medicinal rice of Kerala, Njavara; black/purple landraces of northeast India, Mamihunger (anthocyanin: 96.71 mg/100 g; gammaoryzanols: 0.98 mg/g), Manipuri black (anthocyanin: 96.22 mg/100 g); Kala namak rice of Uttar Pradesh with the highest phenolic content (43.19 ± 0.54 mg/g) and total flavonoid content of 7.18 ± 0.52 mg/g; Mapillai samba of Tamilnadu with anthocyanin content 42.21 mg/g; Chakhao, an aromatic landrace from northeast India, as a rich source of anthocyanins: 275.8 mg/100 g, phenolics: 700 mg GAE/100 g, DPPH antioxidant activity: 65.7%; Kalobhat from West Bengal; Kola bora, a rice landrace of Assam with high anthocyanin content of 328.26 mg/100 g are a rich source of phenolics and flavonoids (Isaac et al. 2012; Sanghamitra et al. 2017, 2018; Rajendran and Chandran 2020; Bhuvaneswari et al. 2020; Kalita and Hazarika 2022).

Many high-yielding pigmented rice varieties enriched with phenolics and flavonoids are black rice cultivars of Japan, Okuno-Murasaki and Asamurasaki; Kuang-fu-Shiang-waxy (KFSW, a waxy indica rice with red-colored bran) and Taikeng 16 (TK16, non-waxy japonica rice) of Taiwan; Thailand aromatic rice, "Rice berry," a deep purple indica-type rice variety; the red rice Rubi and the black rice, Onix of Brazil; black rice genotypes of India with high phenolic content, antioxidant activity, and flavonoids are BPT 3140 (214.34 ± 0.25, 108.83 ± 0.28, 590.12 ± 7.0, respectively), BPT 2848 (23.31 ± 0.26, 86.63 ± 0.65, 784.54 ± 21.6 respectively), BPT 2858 (174.48 ± 0.21, 89.07 ± 0.56, 548.72 ± 7.6, respectively), BPT 3111 (95.50 ± 0.17, 110.09 ± 0.18, 410.61 ± 8.9, respectively) (Mbanjo et al. 2020; Lin and Lai 2011; Takita et al. 2001; Wickert et al. 2014; Sridevi et al. 2021). Chinese wild rice of the tertiary gene pool, *Zizania latifolia,* with higher phenolic content and antioxidant activity, was reported as a potential source of natural antioxidants (Yu et al. 2021).

3.1.8 Rice for Medicinal Use

Certain pigmented rice with medicinal value are used to treat skin diseases, blood pressure, fever, paralysis, rheumatism, and leucorrhea, and even as the basis of a general health tonic (Ahuja et al. 2007). The rice landraces of Madhya Pradesh, such as Gathuan, are used for the treatment of rheumatism; Laichais is used to prevent eponymous skin diseases; Karhani for the treatment of paralysis; Nagkesar for the treatment of lung disease; Resairi for the treatment of chronic cough; and Rakthasali (a kind of red rice) is efficient in subduing disturbed tumors of the body and good for fevers and ulcers; improves eyesight, health, voice and skin health; and increases fertility (Das and Qudhia 2001; Bhat and Riar 2015; Ahuja et al. 2008).

Najvara rice from Kerala is used to treat cervical spondylitis, paralysis, rheumatoid arthritis, neuromuscular disorders, psoriasis, skin lesions, reduce backache, stomach ulcers, and snakebite; and Nivara rice is also used in the preparation of weaning food for underweight babies (Ahuja et al. 2008). In Himachal Pradesh, red rice landraces Matali and Lal Dhan are used for curing blood pressure and fever (Chaudhari et al. 2018). Black rice variety Kalanamak enriched with potassium is used to cure skin diseases and regulate blood pressure. Karuthakar is used to maintain the sugar level under control and cure piles on regular consumption; Kattuyanam rice controls diabetics. Kodai samba rice cures Vatha-related diseases. Karun kuruvai enriched with protein, fat, and phosphorus is used to cure skin ailments and venomous bites and stings (Chaudhari et al. 2018; Pushpam et al. 2019).

The purple pericarp is governed by two genes: purple pericarp b (Pb synonym Prp-b on chromosome 4) and purple pericarp p (Pp, synonym Prp-a on chromosome 1). These genes also show epistatic effects. Dark purple seeds were observed with Pb_PpPp, medium purple seeds with Pb_Pppp, brown seeds with Pb_pppp, and white pericarp seeds with pbpbpppp combination (Mackon et al. 2021).

It is believed that less than 1% of germplasm resources have been exploited by several thousands of accessions. Parallel germplasm screening for all or targeted nutraceuticals should be a continuous process to identify better donors and new genomic regions through genome-wide association studies (GWAS), metabolic pathways, transcription factors, etc.

3.2 QTL Mapping

As most of the grain nutrient traits are controlled by several genes/QTLs, interventions from novel and recent technologies would expedite the development of biofortified rice varieties. The breeding strategies could be fine mapping, marker-assisted selection/backcrossing, association mapping analyses, and natural allelic variation (Priyadarshi et al. 2021). Using bi-parental mapping populations, 22 independent studies have reported 220 QTLs for grain Fe and Zn in rice using simple sequence repeat (SSR) markers or candidate gene-based markers (Raza et al. 2019) (Table 9). Interactions among the identified QTLs for grain Zn were also studied for characterization of QTL interaction effects on the phenotypic expression of the trait

Table 9 QTLs for grain nutrient traits (amino acids, protein, grain Fe, and Zn content) of rice

S. No	Cross/parents	Traits	No. of QTLs	References
1	Zhenshan 97/Nanyangzhan	Amino acid content	2 QTL clusters	Wang et al. 2008
2	Zhenshan 97/Minghui 63	Amino acid content	10 (His) + 8 (Arg)	Zheng et al. 2008
3	Zhenshan 97/Minghui 63	Amino acid content	12	Lu et al. 2009
4	Zhenshan 97B/Delong 208	Amino acid content	3 QTL clusters	Zhong et al. 2011
5	Dasanbyeo/TR22183	Amino acid content	6	Yoo 2017
6	Zhenshan 97/Minghui 63	Protein content	2	Tan et al. 2001
7	Caiapo/*O. glaberrima* (IRGC 103544)	Protein content	4	Aluko et al. 2004
8	Gui630/02428	Protein content	5	Hu et al. 2004
9	V20A/*O. glaberrima* (103,544)	Protein content	1	Li et al. 2004
10	Moritawase/Koshihikari	Protein content	3	Wada et al. 2006
11	Koshihikari/Kasalath// Koshihikari	Protein content	2	Takeuchi et al. 2007
12	Xieqingzao B/Milyang 46	Protein content	5	Lou et al. 2009
13	Zhenshan 97/Minghui 63	Protein content	9	Yu et al. 2009
14	Samgang/Nagdong	Protein content	3	Shi et al. 2009
15	Asominori/IR24	Protein content	9	Qin et al. 2009
16	Asominori/IR24	Protein content	10	Zheng et al. 2011
17	Zhenshan 97B/Delong 208	Protein content	2	Zhong et al. 2011
18	Koshihikari/Kasalath// Koshihikari	Protein content	4	Zheng et al. 2012
19	Chuan/Nanyangzhan	Protein content	2	Zhang et al. 2011
20	Cheongcheong/Nagdong	Protein content	1	Lee et al. 2014
21	CJ06/TN1	Protein content	1	Leng et al. 2014
22	Cheongcheong/Nagdong	Protein content	3	Yun et al. 2014
23	M201/JY293	Protein content	5	Xu et al. 2015
24	Sasanishiki/Habataki	Protein content	1	Yang et al. 2015
25	Cheongcheong/Nagdong	Protein content	1	Bruno et al. 2017
26	ARC 10075//Naveen	Protein content	3	Chattopadhyay et al. 2019
27	IR64/Azucena	Grain Fe and Zn	Fe-3; Zn-2	Stangoulis et al. 2007
28	Zhengshan 97/Minghui 63	Grain Fe and Zn	Fe-2; Zn-3	Lu et al. 2008
29	Teqing/*O. rufipogon* Griff.	Grain Fe and Zn	Fe-1; Zn-2; Fe-1	Garcia-Oliveira et al. 2009
30	Bala/Azucena	Grain Fe and Zn	Fe-4; Zn-4	Norton et al. 2010
31	ZYQ8/JX17	Grain Fe and Zn	Zn-2	Zhang et al. 2011
32	Madhukar/Swarna	Grain Fe and Zn	Fe-7; Zn-6	Anuradha et al. 2012
33	Chunjiang 06/TN1	Grain Fe and Zn	Fe-3; Zn-6	Du et al. 2013

(continued)

Table 9 (continued)

S. No	Cross/parents	Traits	No. of QTLs	References
34	PAU201/Palman 579	Grain Fe and Zn	Fe-8; Zn-3	Kumar et al. 2014
35	Ce258/IR75862 ZGX1/ IR75862	Grain Fe and Zn	Fe-1; Zn-4	Xu et al. 2015
36	Swarna/Moroberekan	Grain Fe and Zn	Fe-1	Indurkar et al. 2015
37	XB/*O. rufipogon* (accession of DWR)	Grain Fe and Zn	Fe-3; Zn-6	Hu et al. 2016
38	Nipponbare/*O. meridionalis* (W1627)//Nipponbare	Grain Fe and Zn	Zn-4	Ishikawa et al. 2017
39	PSBRc82/Joryeongbyeo and PSBRc82 × IR69428	Grain Fe and Zn	Fe-1; Zn-8	Swamy et al. 2018b
40	IR64/IR69428 and BR29/ IR75862	Grain Fe and Zn	Zn-8	Descalsota-Empleo et al. 2019
41	PAU201/Palman	Grain Fe and Zn	Fe-5; Zn-1	Kumar et al. 2019
42	23 mapping populations	Grain Fe and Zn	48 meta QTLs (Fe and Zn-47; Zn-1)	Raza et al. 2019
43	RP Bio 226/Sampada	Grain Fe and Zn	Fe-4; Zn-3	Dixit et al. 2019
44	IR05F102/IR69428	Grain Fe and Zn	Fe-2; Zn-4	Calayugan et al. 2020
45	MTU1010/BR2655	Grain Fe and Zn	Zn-1	Rathod et al. 2021
46	PR116/Ranbir Basmati	Grain Fe and Zn	Fe-5; Zn-2	Suman et al. 2021

(Calayugan et al. 2020; Descalsota-Empleo et al. 2019; Swamy et al. 2018a, b; Jang et al. 2020; Pippal et al. 2022).

3.3 Nutritional Improvement in Rice Quality Using a Genetic Engineering Approach

A deficiency of vitamin A is widespread in countries with cereals as a staple food (Garg et al. 2021). As a natural variation of β-carotene in rice grain is unavailable, the provitamin A pathway into the rice endosperm was introduced through genetic engineering. Several international funding agencies like the Rockefeller Foundation, the Bill & Melinda Gates Foundation, USAID, the Philippine Department of Agriculture, HarvestPlus, the European Commission, Swiss Federal Funding, and the Syngenta Foundation have funded the project of golden rice. Radio-labeled C^{14} indicates that enzymes required for the conversion of isopentenyl-pyrophosphate to geranyl-geranyl-pyrophosphate are available in rice. However, phytoene synthase (*psy*) is unavailable; hence, *daffodil* PSY was introduced with the host (rice) or tissue-specific glutelin promoter, producing a higher amount of phytoene than the *CaMV-35S* promoter. A bacterial *CrtI* gene was chosen over plant sources to avoid

multiple enzyme requirements to convert phytoene to lycopene (Potrykus 2001). However, considering the lower levels of β-carotene, orthologs of PSY were explored, and among them, maize (*ZmPSY1*) was the best, followed by rice (*OsPSY1*). Change in aspartate to alanine in *PSY2* increased the root carotenoid content (Welsch et al. 2010). Provitamin-A noted a half-life of 25 days after harvest, followed by a plateau (3–5 μg/g) of bound (protein) form in the chromoplast (Schaub et al. 2017), and this bound form may be enhanced to increase its quantity and stability (Welsch and Li 2022). Provitamin-A-containing variety (BRRI Dhan 69) was released in Bangladesh. Interestingly, a native provitamin A (49.16 ± 5 μg) pathway was detected in a purple rice (Chettry et al. 2019). Through *Agrobacterium*-mediated transformation, the genetic construct consisted of a plant phytoene synthase (*psy*) originating from daffodil (*Narcissus pseudonarcissus*), a bacterial phytoene desaturase (*crtI*) originating from *Pantoea ananatis* and lycopene ≈-cyclase (*lcy*) from daffodil was introduced into a japonica rice variety as a proof of concept. Second-generation golden rice (GR2) with higher β-carotene levels up to 37 μg/g carotenoids, GR2, was developed using the bacterial phytoene desaturase (*crtI*) originating from *P. ananatis* and phytoene synthase of maize (Paine et al. 2005). Using the backcross breeding program, popular rice varieties of South Asia and Southeast Asian countries, viz., IR64, PSBRc82, and BRRI Dhan29, were introgressed with transgenes associated with golden rice through the backcross breeding approach (www.goldenrice.org). Agronomic evaluation of transgenic and non-transgenic BRRI Dhan 29 showed promising performance for yield (Biswas et al. 2021). Studies have assessed golden rice to be safe as per food safety approvals in various countries (Owens 2018; Oliva et al. 2020). In 2021, golden rice was approved for commercial cultivation in the Philippines (www.goldenrice.org). Regulatory constraints for transgenic rice in several countries are slowing the adoption of transgenic biofortified genotypes.

Transgenic rice lines with high-grain protein content are MH86 (12.87%); Nipponbare line with 6.0% elevation in lysine content; pSBM-35DHPSM rice lines with 2.5- to 4.0-fold increase in the free lysine in the seeds; Kitaake lines with 56% enhancement in lysine level; japonica cv. 9983 with 35% enhancement in lysine; Chinese transgenic male sterile line; Peiai64S (PA64S), with 30% more lysine compared to wild type; another transgenic line of Nipponbare with increased lysine (33.87%), threonine (21.21%), total amino acids (19.43%), and crude protein (20.45%); Taipei 309 line (IRGC accession 42576) with 2.4, 2.0, 1. 4, and 4.8 fold increased cysteine, glutathione, free methionine, and methionine respectively compared to wild type (Zheng et al. 1995; Lee et al. 2001; Kawakatsu et al. 2010; Wong et al. 2015; Liu et al. 2016; Jiang et al. 2016; Xu et al. 2017; Nguyen et al. 2012).

Transgenic rice lines with enhanced vitamin E are WY3 rice lines transformed with *Arabidopsis* γ-tocopherol methyl transferase (γ-TMT gene/AtTMT) driven by the constitutive maize ubiquitin (Ubi) promoter, ASD16 Indica rice lines transformed with two genes from *Arabidopsis thaliana* [tocopherol cyclase (TC) and homogentisate phytyltransferase (HPT)] involved in tocopherol biosynthesis pathway (Zhang et al. 2013; Sundararajan et al. 2021).

Transgenic vitamin B9 or folate rice with increased folate content are Kitaake (14.5–27.2%) (Dong et al. 2016), second-generation folate-enriched transgenic lines GA9.15 and GA26.5 of japonica rice. Nipponbare with improved folate stability was developed by following folylpolyglutamate synthetase (FPGS) strategy that allows folate polyglutamylation that contains 500 μg of folates per 100 g cooked rice that could satisfy the requirement of an average adult (300–400 μg) in three servings of 100 g cooked rice (Blancquaert et al. 2015).

Transgenic rice with enhanced vitamin B1 was developed from the Japonica variety, Kitaake. transformed with two thiamin biosynthesis genes, 4-methyl-5-β-hydroxyethyl thiazole phosphate synthase (THI) and 4-amino-2-methyl-5-hydroxymethyl pyrimidine phosphate synthase (THIC), under the control of the constitutive ubiquitin promoter having fivefold enhanced thiamine in the unpolished, brown rice grain (Dong et al. 2016). Another *japonica* rice Taipei 309 (TP309) has enhanced vitamin B6 (3.1-fold) transformed with two *Arabidopsis thaliana* genes from the vitamin B6 biosynthesis de novo pathway, AtPDX1.1, and AtPDX2, under the constitutive CaMV 35S promoter (Mangel et al. 2019).

Many transgenic lines with enhanced mineral content are transgenic Fe rice (Table 10). Kitaake was transformed with soybean ferritin, SoyferH-1 gene, under the control of the rice seed storage glutelin gene promoter, *GluB-1*, and the rice seed storage globulin gene promoter, Glb-1 with 3.0-fold Fe and 1.1-fold increase in Zn; *OsNAS3-D1* transgenic rice plants was transformed with nicotinamine synthase gene (*OsNAS3*) driven by 35S promoter, with enhanced Fe (2.9-fold), Zn (2.2-fold), and Cu (1.7-fold); *japonica* cv. Taipei 309 transformed with *Arabidopsis thaliana* nicotinamine synthase (*AtNAS1*), *Phaseolus vulgaris* ferritin (*Pvferritin),* and *Aspergillus fumigates* phytase (*Afphytase)* with 6.3-fold Fe and 1.6-fold Zn. Tsukinohikari was transformed with yellow strip-like gene (*OsYSL2*) driven by the sucrose transporter promoter (*OsSUT1*), with 4.4-fold Fe; the japonica variety 'Xiushui 110' was transformed with the rice nicotianamine synthase gene (*OsNAS1*) fused to a rice glutelin promoter (*GluB1*) with 3.0-fold Fe and 2.7-fold Zn; the "Paw San Yin" Myanmar transgenic variety was transformed with nicotianamine synthase gene HvNAS1 driven by 35S promoter, nicotianamine transporter gene *OsYSL2,* and Fe storage protein gene *SoyferH2* driven by endosperm-specific globulin promoter with 3.4-fold Fe and 1.3-fold Zn increase compared to the wild type (Lee et al. 2009a, b; Qu et al. 2005; Wirth et al. 2009; Ishimaru et al. 2010; Zheng et al. 2010; Aung et al. 2013).

Transgenic lpa rice lines are Kitaake lines with 68% reduction in the PA by suppression of RINO1 gene expression driven by the glutelin B-1 (GluB-1) and 18-kDa oleosin 18 (Ole 18) promoters (Kuwano et al. 2009); PusaSugandhi II (IO6-97) with 69% reduction in PA level developed by silencing the IPK1 gene (inositol 1,3,4,5,6-pentakisphosphate 2-kinase) using the Ole 18 promoter in an RNAi-mediated approach (Ali et al. 2013). Wild rice genotypes (*Oryza rufipogon*) Lua Vang and Lua Hoang of the primary gene pool were reported having a low phytic acid content of 0.930 μg P and 1.395 μg P, respectively (Lang et al. 2007).

Genetically engineered high-resistant starch lines are indica rice variety, Teqing-resistant starch rice line (TRS) enriched with high amylose (60–64.8%) developed

Table 10 Transgenic/genetic engineering approaches for rice biofortification

S. No	Rice cultivars	Approaches	Promoter and genes targeted	Range of fold increase over wild	Reference
1.	*Indica* cv. IR64 *Indica* cv. IR68144 *Indica* cv. Kasalath *Japonica* cv. Dongjin *Japonica* cv. Kitaake *Japonica* cv. Nipponbare *Japonica* cv. Taipei 309 *Japonica* cv. Tsukino Hikari *Japonica* cv. Xiushui 110 *Japonica* cv. Zhonghua 10 *Japonica* cv. Zhonghua 11 *O. sativa* cv. EYI 105 *O. sativa* cv. EYI 105	Approaches involving transporters and Fe sequestration; manipulation of the phytosiderophore pathway; changing the expression of metal homeostasis-related genes (Fe homeostasis regulators, transporters, and binding molecules) and other approaches that resulted in increased Fe concentration	*Activation tag line of OsNAS2*; *Activation tag line of OsNAS3*; *CaMV 35S::MxIRT1*; *CaMV35S::AtCOPT1*; *CaMV35S::HvTOM1*; *CaMV35S::OsbHLH058; CaMV35S::OsIMA1; CaMV35S::OsIMA2; CaMV35S::OsIRO2*; *CaMV35S::OsNAS1; CaMV35S::OsNAS2; CaMV35S::OsNAS3; CaMV35S::OsRMC; osrmci; CaMV35S:: OsTOM1*; *CaMV35S::OsYSL13; CtCBM-IBP; Knockout lines by CRISPR/Cas9 osvmt; OsActin::OsYSL15; OsActin1::HvNAS1; OsbHLH059i; OsGlu::AfPhytase; OsGluB1::GmFer; OsGluB1:: GmFerH1; OsGluB1::OsNAS1; OsGluB4::GmFerH1; OsHRZ1i + OsHRZ2i; Osmit; OsSUT1::OsYSL2; OsUbi::OsRab6a; osysl9i; point mutation osnaat1; T-DNA insertion mutant line:osvit1; T-DNA insertion mutant line:osvit2; Ubi::OHMA7 allele 261; Ubi::OHMA7 allele 284; ZmUbi::HvYS1; ZmUbi::OsIRT1; ZmUbi::OsNAS1 + ZmUbi::HvNAATb; ZmUbi::OsNAS2*	Fe (0.83–8.20 fold)	Goto et al. 1999 Yang et al. 2020 Yang et al. 2013 Andrés-Bordería et al. 2017 Bashir et al. 2013 Banakar et al. 2017a Zhang et al. 2018 Senoura et al. 2017 Ishimaru et al. 2010 Lee et al. 2009a Che et al. 2019 Zhang et al. 2012 Bashir et al. 2013 Zhang et al. 2012 Kappara et al. 2018 Kobayashi et al. 2019 Kobayashi et al. 2019 Ogo et al. 2011 Kobayashi et al. 2021 Kobayashi et al. 2013 Nozoye et al. 2011 Díaz-Benito et al. 2018 Banakar et al. 2017b Cheng et al. 2007b Zheng et al. 2010 Masuda et al. 2009 Johnson et al. 2011 Lee et al. 2012

(continued)

Table 10 (continued)

S. No	Rice cultivars	Approaches	Promoter and genes targeted	Range of fold increase over wild	Reference
					Lee et al. 2009b Oliva et al. 2014 Vasconcelos et al. 2003 Lucca et al. 2001 Tan et al. 2015 Lee and An 2009
2.	*Japonica* cv. Dongjin *Japonica* cv. Nipponbare	Zn biofortification-still lacking basic knowledge	*CaMV35S::OsZIP9; osmt2b; osmt2c; osmt2bosmt2c; dmas*	Zn (0.8–1.3-fold)	Bashir et al. 2017; Tan et al. 2020; Lei et al. 2021
3.	*Indica* cv. IR64 Tropical Jap*onica* cv. Paw San Yin *Japonica* cv. Tsukino Hikari *Japonica* cv. Taipei 309 *Japonica* cv. Nipponbare	Combination of approaches	*CaMV35S::OsNAS2 + GluA2::GmFer-H1; OsGluB1::GmFer-H2 + OsGluB1::OsYSL2 + OsGluB1::GmFer-H2 + OsActin1:: HvNAS1 + OsSUT1::OsYSL2; OsGlb::GmFerH2 + OsGluB::: HvNAAT-A + OsGluB::HvNAS1 + OsGluB1::GmFerH2 + OsGluB1:: HvIDS3; OsGlB::PvFer + OsGlB::AfPhytase + CaMV35S::AtNAS; MsEnod12B::AtIRT1; MsEnod12B::AtIRT1 + OsGluB1::PvFER + aMV35S::OsNAS1 AtIRT1::AtIRT1 + OsGluB1:: PvFER + CaMV35S::OsNAS1; OsGlB1::PvFER + CaMV35S:: AtNAS1 + ZmUbi::AtFRD3 OsGlB1::PvFER + ZmUbi:: AtFRD3; CaMV35S:: AtNAS1 + OsGlB1::PvFER + ZmUbi::AtNRAMP3 CaMV35S::AtNAS1 + OsGlB1:: PvFER + OsOle18::AtNRAMP3 OsGlB1::PvFER + ZmUbi:: AtNRAMP3; OsGlB1::PvFER +OsOle18::AtNRAMP3; OsHMA2::ZmYS1 OsFRDL1:OsTOM1; OsHMA2::ZmYS1 + OsFRDL1:OsTOM1; OsHMA2::ZmYS1 + OsFRDL1: OsTOM1 + OsGlob1:GmFer H1; OsHMA2::ZmYS1 + OsFRDL1: OsTOM1 + OsGlob1:GmFer H1 +ZmUbq1::HvNAS1*	Fe (1.8–9.30-fold) Zn (1.2–3.8-fold)	Wirth et al. 2009 Masuda et al. 2013 Aung et al. 2013 Trijatmiko et al. 2016 Kawakami et al. 2022 Wu et al. 2019 Wu et al. 2018 Boonyaves et al. 2017 Boonyaves et al. 2016

by Antisense RNA inhibition of starch branching enzymes, SBEIIb and SBEI (Zhu et al. 2012). Genome-edited high-resistant starch rice line is *Japonica* cv. Kitaakelines were developed through CRISPR/Cas9-mediated target gene editing of starch branching enzymes, SBEIIb, and SBEI, with a 9.8% increase in resistant starch (Sun et al. 2017).

Genetic engineered rice varieties enriched with phenolics and flavonoids are the transgenic line, Hwa-Young, transformed with maize *C1/R-S* regulatory genes using the promoter of a rice prolamin gene produces numerous kinds of flavonoid compounds in the endosperm; Taichung 65, transformed with regulatory gene OSB2, encodes a transcription factor for anthocyanin synthesis developed by using OSPR 1.1 promoter; GluBpro-Naringenin and Olepro-Naringenin transgenic rice with enhanced flavonoids content were developed by overexpression of maize C1 and R-S regulatory genes [Myb-type and basic helix-loop-helix-type transcription factors:phenylalanine ammonia-lyase and chalcone synthase (CHS) genes]; lines of white rice Nipponbare and Taichung 65 were transformed with Myc-type bHLH gene *OsB2* from the black rice variety Khum by using *Agrobacterium tumefaciens* with enhanced anthocyanin synthesis (Shin et al. 2006; Kawahigashi et al. 2007; Ogo et al. 2011; Sakulsingharoj et al. 2014).

Recently genetically engineered 'Purple Endosperm Rice' (Zijingmi) were developed from the *japonica* variety Zhonghua 11 (ZH11) and the *indica* variety Huaguang 1 (HG1) and transformed with a transgene stacking system (TS II) containing two regulatory genes (maize *leaf color, ZmLc,* which encode the bHLH-type and maize purple leaf, *ZmPl,* which encode the MYB-type transcription factors, that activate the anthocyanin biosynthesis genes) and six structural anthocyanin-related genes from *Coleus*. These are driven by eight endosperm-specific promoters (ESPs) from rice. It has high anthocyanin content and antioxidant activity in the endosperm (Zhu et al. 2017).

p-Coumaroylserotonin and feruloyl serotonin polyphenol compounds are two important serotonin derivatives known for antioxidative activity. Transgenic rice with higher radical scavenging activities express pepper hydroxycinnamoyl-CoA: serotonin N-(hydroxycinnamoyl) transferase gene. This transgenic rice was observed with a serotonin content of 274 ng/g seed weight ninefold higher than wild type (30 ng/g seed weight) (Kang et al. 2005).

Coenzyme Q (CoQ), or ubiquinone, is a lipid-soluble antioxidant. It is also an electron transfer molecule in the respiratory chain. Most cereal crops produce mainly CoQ9, which has nine isoprene units. At the same time, CoQ10 is the typical food supplement human beings use. Transgenic rice CoQ10 (Nipponbare) introgressed with decaprenyl diphosphate synthase is responsible for synthesizing 10 isoprene units for the CoQ side chain from Gluconobacter suboxydans 3 are able to produce 1.3–1.6 times more CoQ10. This gene was transferred into rice sugary and shrunken mutants of Nipponbare deficient in starch biosynthesis and had a wrinkled seed (Takahashi et al. 2010).

During rice domestication, wild rice species having pigmented pericarp (Rc) were deselected for nonpigmented pericarp rice having an independently evolved loss-of-function mutation *rc* (C → A substitution in exon 7). However, in US and Italian

cultivars, the two reversal (rc to Rc) mutations are found, 1-bp deletion located 20-bp upstream and 1-bp deletion located 44-bp upstream of the original 14-bp deletion, respectively, and many more might be present in the cultivars available across the world. Sequence analysis of the 4 MB region around the Rc in 156 rice indicates that the reverse mutations are not responsible for pigmentation in weedy rice, which contains haplotypes that are either rare or absent in cultivated rice (Gross et al. 2011). Therefore, it appears reasonable to sequence this region in all accessions or core sets of pigmented rice to understand this phenomenon.

A multidrug resistance-associated protein (*Os*MRP) activates a membrane-bound protein (*Os*MRP15) that aids in anthocyanin uptake into the vacuole. The introduction of nonsense mutation in OsMRP15 gene (CRISPR/Cas 9) resulted in the formation of green leaves in the mutant, while the wild-type leaves were purple (Ma et al. 2015). The expression of anthocyanin synthesis genes is activated by binding the MYB (myeloblastosis) proteins to DNA. Rice has multiple gene copies (paralogues) responsible for production in different patterns or cell types. Many transgenic rice varieties show purple and black endosperm, having good nutritional and medical values that are useful to the nutraceutical and pharmaceutical industries.

4 Genomics-Aided Breeding for Health-Related Traits

With recent advances in high-throughput sequencing, phenotyping platforms, and genotyping technologies, molecular breeding has been transformed into genomics-aided breeding (Kole et al. 2015). Both classical genetics and NGS platforms, a sequencing-based trait mapping, aid in identifying/mapping nutrition traits at a higher resolution (Sinha et al. 2021). Genome-wide association analyses have elucidated the genetic architecture, marker–trait association, and haplotypes in different populations and germplasm panels (Sahu et al. 2020). Genetic analysis of both GWAS and genomic selection (GS) helps precise nutritional enrichment in rice grain (Zaghum et al. 2022). A core collection of 3000 rice accessions (3K) has been screened for large-scale discovery of novel alleles for different traits by applying various bioinformatics and genetic approaches. Key genes influencing Fe and Zn have been surveyed in the 3K rice genome panel (Abbai et al. 2019). Implementing advanced genomics, transcriptomics, nutrigenomics, and proteomics-based technologies enhances and regulates the nutritional value of rice.

4.1 GWAS

In rice, the GWAS approach has been demonstrated to identify several genes and genomic regions associated with important agro-morphological traits (Huang et al. 2012, 2015). GWAS is based on genetically diverse genotypes with several historic recombination events and high-throughput markers like single-nucleotide polymorphisms (Crowell et al. 2016; Yano et al. 2016). Using SNP markers, genomic regions for grain nutrients and bioactive compounds have also been identified in rice using

genome-wide association studies (GWAS) (Norton et al. 2014, Zhang et al. 2014; Pradhan et al. 2020; Bollinedi et al. 2020; Cu et al. 2021; Alpuerto et al. 2022) and also in multi-parent advanced generation inter-cross (MAGIC) population (Descalsota et al. 2018).

4.2 Sequencing/Resequencing

After completing the International Rice Genome Sequencing Project, several rice genotypes have been sequenced and assembled, resulting in high-quality genome sequences in rice with the advent of high-throughput short-read and long-read sequencing technologies (IRGSP 2005). Highly continuous and accurate chromosome-level assemblies with minimum gaps are made possible by deploying ultra-long reads and new assembly tools. Using low-cost sequencing technologies leading to resequencing of the popular donors and functional genomics approach, candidate genes associated with nutrient metabolism are being identified (Wang and Han 2022).

4.3 Wild Species

Wild rice species is a valuable resource of several agro-morphological traits such as yield, quality, and resistance to biotic and abiotic stresses (Fan et al. 2019). Identifying new genes or genomic regions associated with nutrients and antinutrients in wild species appears promising, as reported in *Oryza rufipogon* and *Oryza longistaminata* (Garcia-Oliveira et al. 2009; Anuradha et al. 2012; Hu et al. 2016; Liu et al. 2020). Recent genome sequencing projects involving wild rice species have revealed enormous sequence and allele diversity that can be fully exploited and utilized in rice improvement programs (Jacquemin et al. 2013; Mussurova et al. 2020). Marker-delimited genomic region for Fe and Zn contents from *O. longistaminata* could be a basis for subsequent gene cloning (Liu et al. 2020).

4.4 3K Rice Panel

A germplasm panel constituting 3010 genotypes (3K) from the core collection of Asian cultivated rice germplasm from 89 countries was resequenced by the International Rice Research Institute (IRRI) and Chinese Academy of Agriculture Science (Wang et al. 2018). This resource is now being phenotyped worldwide for various traits toward identifying genes/genomic regions for traits of interest. Genotype data is available as 29 million single-nucleotide polymorphisms (SNPs), around 2.4 million small insertions and deletions (indels), and >90,000 structural variants for 3K with 23,876 gene families, 14,826 core genes, and 9050 species-specific genes/gene families (Sun et al. 2017). The most exciting aspect was the 12,465 novel genes identified in the 3K panel, which were not present in the Nipponbare reference

genome. 3K has also been characterized for gene–coding sequence–haplotype (gcHap) diversity of 45,963 rice genes (Zhang et al. 2021). The 3K panel has been preliminarily surveyed for nutrient-related genes in rice (Abbai et al. 2019).

4.5 Genomic Selection

The methodology of GS comprises the prediction of genome-estimated breeding values from a training population with known genotype and phenotype and the prediction of the phenotype of the test population without the genotype information (Crossa et al. 2017). The GS strategy can be applied in rice breeding as genotype information is available for several germplasm panels and has been successfully demonstrated in hybrid prediction, grain filling, and other agro-morphological traits (Xu et al. 2018). According to the breeder's equation, even a slight increase in predictive ability can be converted into a massive gain with a vigorous selection intensity (Xu et al. 2021b). By deploying a synthetic population comprising upland rice cultivars and evaluating at two locations, the early genomic prediction has reported a predictive ability of 0.51 for grain Zn concentration (Baertschi et al. 2021).

4.6 Functional Genomics

While the structural genomics of rice has been deciphered through genome sequencing, functional genomics of identified genes is essential to validate the function and characterization of identified genes. Many genes have been characterized by the generation of full-length cDNA libraries, gene expression microarrays, and RNA sequencing technologies (Li et al. 2018). Two cDNA databases of *japonica*, Knowledge-Based *Oryza* Molecular Biological Encyclopedia (KOME) and *indica* as Rice indica cDNA Database (RICD), are made available. The specific expression profiles of various tissues for hybrid and its parental lines are also being deposited in the collection of the Rice Expression Profiles (CREP) database. Another database that rice researchers extensively use is the Rice Expression Profile Database (RiceXpro), which is a repository of gene expression profiles derived from microarray analysis of tissues and organs across the entire growth of the rice plant. RNA sequencing is another methodology that is frequently applied to identify tissue, stage, and situation-specific gene expression that is being deployed to characterize various genes associated with grain nutrient metabolism (Neeraja et al. 2018; Kar et al. 2021; Li et al. 2020). Gene expression databases constituting functionally related genes in various biological pathways and metabolic processes are also available.

4.7 Bioinformatics

With the application of various next-generation sequencing (NGS) technologies, several public databases are made available for genome sequences, transcriptomes,

proteomes, metabolomes, and phenotype traits. Several bioinformatics websites are available for genome annotation, gene expression, gene function, and interaction (Li et al. 2018; Jia et al. 2021; Chen et al. 2022). Recent bioinformatics databases in rice also encompass information about both functional and regulatory elements. A few comprehensive databases include information about genotypes, genome sequence and phenotypes such as SNP-Seek, RiceVarMap, and Rice Pan-genome Browser (RPAN) (Zhao et al. 2015; Sun et al. 2017; Mansueto et al. 2017).

4.8 Whole-Genome Selection and Breeding Chips

Based on publicly available genome information, several SNP chips ranging from low to high resolution have been developed for genetic diversity analyses, QTL identification, genome-wide association studies, marker-assisted selection, and genomic selection such as RICE6K, with 5102 SNPs and indels, with an average spacing of 6.4 kb between adjacent SNPs (Yu et al. 2014; Thomson et al. 2017); RICE60K (also known as RiceSNP50) with 43,386 SNPs and RICE90K with 85,000 SNPs (Chen et al. 2014; Qiu et al. 2018). Another 50K chip with 50,051 SNPs from 18,980 different genes spanning 12 rice chromosomes, including 3710 single-copy (SC) genes conserved between wheat and rice, 14,959 SC genes unique to rice, 194 agronomically essential cloned rice genes, and 117 multi-copy rice genes were also made available (Singh et al. 2015). A 700 K rice and 44K array chips were also developed mainly for GWAS (McCouch et al. 2016; Zhao et al. 2011). Recently 1k-RiCA chip with 995 SNPs with an average 1.53 cM marker distance was also made available to be deployed by breeders (Arbelaez et al. 2019).

4.9 Genomics-Assisted Breeding

Taking a cue from the Green Super Rice strategy, the following methodology is being proposed for developing varieties with high grain nutrient levels (Yu et al. 2022b). Identifying a recipient parent, mostly an elite cultivar with desirable yield levels and quality parameters, and its characterization through whole-genome sequencing is essential. Further identification of a matching donor with target alleles, viz., high grain nutrient followed by the most efficient scheme of crossing, thus generates a series of near-isogenic lines (NILs), which can be directly evaluated for varieties and as parents for stacking genes. Further, designed QTLs pyramiding combining key grain nutrient traits from promising selective-introgression lines (SILs) were derived from different donor lines but with a common recipient parent. These essential nutritional traits could be genetically mapped and, at the same time, since they are already in elite backgrounds, could be quickly released as varieties with superior nutritional value. Further, designed QTL pyramiding of the nutritional traits from different donor backgrounds could be stacked into a common recipient parent by crossing the different SILs in less than 3 years.

5 Role of Genome Editing Technology in Rice Nutritional Quality Improvement

Among several strategies/approaches toward supporting the development of rice biofortified genotypes, CRISPR-based genome editing appears to be the most promising owing to its targeted gene editing (Kumar et al. 2022). Because of its straightforward design, low methodology cost, high efficiency, good reproducibility, and quick cycle, this system is being deployed in several crops like rice, wheat, barley, maize, potato, and tomato. Several genes for grain dimension, viz., *Gs3*, *Gn1a*, *GL2/OsGRF4*, *OsGRF3*, *GL3.1*, *OsGBSSI,* and *Wx,* were targeted for genome editing (Shen et al. 2018; Hao et al. 2019; Zhou et al. 2019; Yuyu et al. 2020; Huang et al. 2020; Achary and Reddy 2021). The aroma was also enhanced through gene editing of the *BADH2* gene (Ashokkumar et al. 2020; Hui et al. 2022).

The genes for (GABA) content were phytic acid and protein content (Akama et al. 2020; Khan et al. 2019; Wang et al. 2020).

The precision and specificity of genome editing can be utilized to engineer metabolic pathways controlling nutraceutical properties of major crop produce. The successful knockout of FAD2-1A, FAD2-1B, and FAD3A genes by TALEN was employed in soybean to enhance the oleic acid content >80% in oil. These genes play a central role in metabolizing the oleic acid (monounsaturated fatty acid) into linoleic acid (polyunsaturated fatty acid) (Demorest et al. 2016). Recently, biofortified tomatoes enriched with vitamin D have been developed using genome editing. 7-Dehydrocholesterol reductase (Sl7-DR2) gene was knocked out using CRISPR/Cas9 for higher accumulation of provitamin D3 in genome-edited tomatoes. In rice also, increased resistant starch and low casein accumulation traits have been attempted by targeting SBEI, SBEIIb, and OsHAK-1 genes through CRISPR/Cas9 approach. These genes regulate amylose content and calcium uptake phenotypes (Nagamine and Ezura 2022). The key regulatory genes regulating grain iron and zinc content in rice can be edited through various approaches of genome editing, including base editing and prime editing, to create superior alleles in popular rice cultivars. This approach can be instrumental in achieving the mission of nutritional security.

6 Conclusion

Rice, the paramount grain for ages, has shaped human civilization and played a critical role in food security. With the increasing population, the concomitant development and release of high-yielding rice cultivars played a significant role in feeding >7.98 billion people on this planet. The past two decades have brought revolutionary changes in breeding technologies. We have witnessed the wise and efficient use of marker-assisted breeding, genomics, transgenics, and genome editing to accelerate breeding efforts and make them more precise and targeted. Several cultivars with added traits of economic importance have been developed using these biotechnology-based breeding approaches. In particular, CRISPR/Cas9 genome

editing technology can address many unsolved breeding questions and can be smartly used to breed new cultivars for present and future needs (Fig. 6).

While food security will always remain a primary goal, nutrition security is equally essential, specifically in the developing world. Special programs by various governments have been implemented to minimize the adverse effects of malnutrition. However, it is still prevalent in places where rice constitutes the central portion of the diet. Therefore, enriching the rice grain with micronutrients and vitamins can address the severe issue of malnutrition. New genomics and genome modification approaches can be used to biofortify rice crops. In addition to discovering new genes and alleles, the known genes and natural genetic variations must be exploited through breeding and genetic engineering to develop biofortified rice varieties. The regulatory genes and transcription factors determining the loading of micronutrients in rice grain are primary targets for making the rice grain more nutritive. In addition, efforts should be made toward reducing the antinutritional factors such as phytate in rice grain. Targeted breeding programs and the introduction of nutrition-rich rice in the food supply chain will play a key role in achieving the goals of "food and nutritional security." To promote the cultivation of biofortified crops, incentives by the government and special prices for farmers' produce will significantly impact the success of these programs. The goal of zero hidden hunger and starvation is achievable with the focused and coordinated efforts of researchers, funding agencies, policymakers, industries, and farmers.

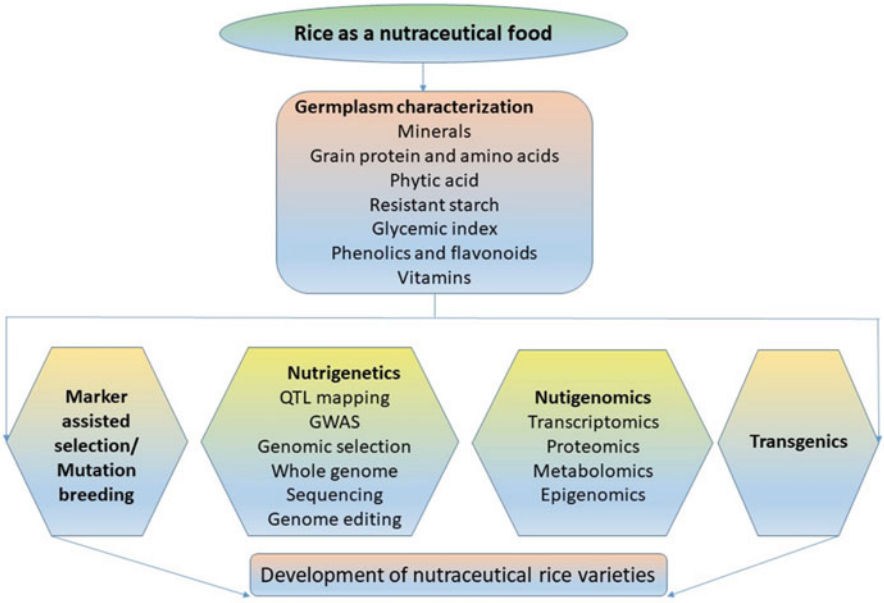


Fig. 6 Genetic strategies toward the development of nutraceutical rice varieties

References

Abbai R, Singh VK, Nachimuthu VV, Sinha P, Selvaraj R, Vipparla AK, Singh AK, Singh UM, Varshney RK, Kumar A (2019) Haplotype analysis of key genes governing grain yield and quality traits across 3K RG panel reveals scope for the development of tailor-made rice with enhanced genetic gains. Plant Biotechnol J 17(8):1612–1622

Abeysiriwardena DS d Z, Gunasekara DCS (2020) Development of a red rice variety with excellent health properties and attractive grain qualities. Indian J Genet 80:115–117

Achary V, Reddy MK (2021) CRISPR-Cas9 mediated mutation in grain width and weight2 (GW2) locus improves aleurone layer and grain nutritional quality in rice. Sci Rep 11(1):1–13

Ahmed MS, Bashar MK, Wazuddin M, Shamsuddin AKM (2015) Morpho-physicochemical characterization of Balam rice (*Oryza sativa* L.) land races of Bangladesh. Int J Plant Breed Genet 9:163–176

Ahuja U, Ahuja S, Chaudhary N, Thakrar R (2007) Red rices – past, present, and future. Asian Agri Hist 11:291–304

Ahuja SC, Thakrar R, Shobha Rani N (2008) Scented rices of India. Asian Agri Hist 12(4):267–283

Akama K, Akter N, Endo H, Kanesaki M, Endo M, Toki S (2020) An in vivo targeted deletion of the calmodulin-binding domain from rice glutamate decarboxylase 3 (OsGAD3) increases γ-aminobutyric acid content in grains. Rice 13(1):1–12

Ali N, Paul S, Gayen D et al (2013) RNAi mediated down regulation of myo-inositol-3-phosphate synthase to generate low phytate rice. Rice 6:12. https://doi.org/10.1186/1939-8433-6-12

Alpuerto JBB, Samonte SOP, Sanchez DL, Croaker PA, Wang YJ, Wilson LT, Christensen EF, Tabien RE, Yan Z, Thomson MJ (2022) Genomic association mapping of apparent amylose and protein concentration in milled rice. Agronomy 12(4):857

Aluko G, Martinez C, Tohme J, Castano C, Bergman C, Oard JH (2004) QTL mapping of grain quality traits from the inter-specific cross between *Oryza sativa* × *Oryza glaberrima*. Theor Appl Genet 109:630

Anacleto R, Badoni S, Parween S et al (2019) Integrating a genome-wide association study with a large-scale transcriptome analysis to predict genetic regions influencing the glycaemic index and texture in rice. Plant Biotechnol J 17:12611275. https://doi.org/10.1111/pbi.13051

Andrés-Bordería A, Andrés F, Garcia-Molina A, Perea-García A, Domingo C, Puig S, Peñarrubia L (2017) Copper and ectopic expression of the Arabidopsis transport protein COPT1 alter iron homeostasis in rice (*Oryza sativa* L.). Plant Mol Biol 95(1):17–32

Anuradha K, Agarwal S, Rao YV, Rao KV, Viraktamath BC, Sarla N (2012) Mapping QTLs and candidate genes for iron and zinc concentrations in unpolished rice of Madhukar× Swarna RILs. Genetics 508:233

Arbelaez JD, Dwiyanti MS, Tandayu E, Llantada K, Jarana A, Ignacio JC, Platten JD, Cobb J, Rutkoski JE, Thomson MJ, Kretzschmar T (2019) 1k-RiCA (1K-Rice Custom Amplicon) a novel genotyping amplicon-based SNP assay for genetics and breeding applications in rice. Rice 12(1):1–15

Asai H, Abe N, Matsushima R et al (2014) Deficiencies in both starch synthase IIIa and branching enzyme IIb lead to a significant increase in amylose in SSIIa-inactive japonica rice seeds. J Exp Bot 65:5497–5507

Aung MS, Masuda H, Kobayashi T, Nakanishi H, Yamakawa T, Nishizawa NK (2013) Iron biofortification of Myanmar rice. Front Plant Sci 4:158

Baek JA, Chung NJ, Choi KC, Hwang JM, Lee JC (2015) Hull extracts from pigmented rice exert antioxidant effects associated with total flavonoid contents and induce apoptosis in human cancer cells. Food Sci Biotechnol 24:241–247

Baertschi C, Cao TV, Bartholomé J, Ospina Y, Quintero C, Frouin J, Bouvet JM, Grenier C (2021) Impact of early genomic prediction for recurrent selection in an upland rice synthetic population. G3 11(12):jkab320. https://doi.org/10.1093/g3journal/jkab320

Bairagi S, Mishra AK, Mottaleb KA (2022) Impacts of the COVID-19 pandemic on food prices: evidence from storable and perishable commodities in India. PLoS One 17(3):e0264355

Banakar R, Alvarez Fernandez A, Abadía J, Capell T, Christou P (2017a) The expression of heterologous Fe (III) phytosiderophore transporter Hv YS 1 in rice increases Fe uptake, translocation and seed loading and excludes heavy metals by selective Fe transport. Plant Biotechnol J 15(4):423–432

Banakar R, Alvarez Fernandez A, Díaz-Benito P, Abadia J, Capell T, Christou P (2017b) Phytosiderophores determine thresholds for iron and zinc accumulation in biofortified rice endosperm while inhibiting the accumulation of cadmium. J Exp Bot 68(17):4983–4995

Bashir K, Takahashi R, Akhtar S, Ishimaru Y, Nakanishi H, Nishizawa NK (2013) The knockdown of OsVIT2 and MIT affects iron localization in rice seed. Rice 6(1):1–7

Bashir K, Nozoye T, Nagasaka S, Rasheed S, Miyauchi N, Seki M, Nakanishi H, Nishizawa NK (2017) Paralogs and mutants show that one DMA synthase functions in iron homeostasis in rice. J Exp Bot 68(7):1785–1795

Berni R, Cantini C, Romi M, Hausman JF, Guerriero G, Cai G (2018) Agrobiotechnology goes wild: ancient local varieties as sources of bioactives. Int J Mol Sci 19(8):2248

Beynen AC, Katan MB (1985) Why do polyunsaturated fatty acids lower serum cholesterol? Am J Clin Nutr 42:560–563. https://doi.org/10.1093/AJCN/42.3.560

Bhat FM, Riar CS (2015) Health benefits of traditional rice varieties of temperate regions. Med Aromat Plants 4:198

Bhuvaneswari S, Gopala Krishnan S, Bollinedi H et al (2020) Genetic architecture and anthocyanin profiling of aromatic rice from Manipur reveals divergence of Chakhaolandraces. Front Genet 11:570731

Biswas PS, Swamy BM, Kader MA, Hossain MA, Boncodin R, Samia M, Hassan ML, Wazuddin M, MacKenzie D, Reinke R (2021) Development and field evaluation of near-isogenic lines of GR2-EBRRI dhan29 golden rice. Front Plant Sci 12:619739

Blancquaert D, Van Daele J, Strobbe S, Kiekens F, Storozhenko S, De Steur H, Glleynck X, Lambert W, Stove C, Van Der Straeten D (2015) Improving folate (vitamin B9) stability in biofortified rice through metabolic engineering. Nat Biotechnol 33:1076–1078

Bollinedi H, Yadav AK, Vinod KK, Gopala Krishnan S, Bhowmick PK, Nagarajan M, Neeraja CN, Ellur RK, Singh AK (2020) Genome-wide association study reveals novel marker-trait associations (MTAs) governing the localization of Fe and Zn in the rice grain. Front Genet 11:213

Bollinedi H, Neeraja CN, Chattopadhyay K, Chandel G, Shashidhar H, Prakash J, Singh AK, Voleti SR, Sundaram RM (2022) Karuppunel: a promising donor for high Zn content in rice (*Oryza sativa*) grain. Indian J Agri Sci 92(10):1247–1252

Boonyaves K, Gruissem W, Bhullar NK (2016) NOD promoter-controlled AtIRT1 expression functions synergistically with NAS and FERRITIN genes to increase iron in rice grains. Plant Mol Biol 90(3):207–215

Boonyaves K, Wu TY, Gruissem W, Bhullar NK (2017) Enhanced grain iron levels in rice expressing an iron-regulated metal transporter, nicotianamine synthase, and ferritin gene cassette. Front Plant Sci 8:130

Bouis HE, Saltzman A (2017) Improving nutrition through biofortification: a review of evidence from HarvestPlus, 2003 through 2016. Glob Food Sec 12:49–58

Bouis HE, Saltzman A, Birol E (2019) Improving nutrition through biofortification. In: Fan S, Yosef S, Pandya-Lorch R (eds) Agriculture for improved nutrition: seizing the momentum. CAB International 2019, pp 47–57

Bruno E, Choi YS, Chung IK, Kim KM (2017) QTLs and analysis of the candidate gene for amylose protein and moisture content in rice (*Oryza sativa* L.). 3 Biotech 7:40

Burrell MM (2003) Starch: the need for improved quality or quantity – an overview. J Exp Bot 54: 451–456. https://doi.org/10.1093/jxb/erg049

Butardo VM Jr, Sreenivasulu N (2016) Tailoring grain storage reserves for a healthier rice diet and its comparative status with other cereals. Int Rev Cell Mol Biol 323:31–70

Butardo VM, Fitzgerald MA, Bird AR et al (2011) Impact of down-regulation of starch branching enzyme IIb in rice by artificial microRNA- and hairpin RNA-mediated RNA silencing. J Exp Bot 62:4927–4941. https://doi.org/10.1093/jxb/err188

Butelli E, Titta L, Giorgio M, Mock HP, Matros A, Peterek S, Schijlen EG, Hall RD, Bovy AG, Luo J (2008) Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nat Biotechnol 26(11):1301–1308

Calayugan MIC, Formantes AK, Amparado A, Descalsota-Empleo GI, Nha CT, Inabangan-Asilo MA, Swe ZM, Hernandez JE, Borromeo TH, Lalusin AG, Mendioro MS (2020) Genetic analysis of agronomic traits and grain Fe and Zn concentrations in a doubled haploid population of rice (*Oryza sativa* L.). Sci Rep 10(1):1–14

Chakuton K, Puangpronp D, Nakornriab M (2012) Phytochemical content and antioxidant activity of colored and non-colored Thai rice cultivars. Asian J Plant Sci 11(6):285–293. https://doi.org/10.3923/ajps.2012.285.293

Chandel G, Premi V, Vishwakarma AK et al (2016) In vitro identification of low glycemic index (GI) white rice using Nutriscan GI Analyzer. Adv Life Sci 5:11090–11098

Chattopadhyay K, Das A, Das SP (2011) Grain protein content and genetic diversity of rice in north eastern India. Oryza 48(1):73–75

Chattopadhyay K, Behera L, Bagchi TB, Sardar SS, Moharana N, Patra NR et al (2019) Detection of stable QTLs for grain protein content in rice (*Oryza sativa* L.) employing high throughput phenotyping and genotyping platforms. Sci Rep 9:3196

Chaudhari PR, Tamrakar N, Singh L et al (2018) Rice nutritional and medicinal properties: a review article. J Pharmacogn Phytochem 7:150–156

Che J, Yokosho K, Yamaji N, Ma JF (2019) A vacuolar phytosiderophore transporter alters iron and zinc accumulation in polished rice grains. Plant Physiol 181(1):276–288

Chen H, Xie W, He H, Yu H, Chen W, Li J, Yu R, Yao Y, Zhang W, He Y, Tang X (2014) A high-density SNP genotyping array for rice biology and molecular breeding. Mol Plant 7(3):541–553

Chen R, Deng Y, Ding Y, Guo J, Qiu J, Wang B, Wang C, Xie Y, Zhang Z, Chen J, Chen L (2022) Rice functional genomics: decades' efforts and roads ahead. Sci China Life Sci 65:33–92

Cheng L, Wang F, Shou H, Huang F, Zheng L, He F, Li J, Zhao FJ, Ueno D, Ma JF, Wu P (2007b) Mutation in nicotianamine aminotransferase stimulated the Fe (II) acquisition system and led to iron accumulation in rice. Plant Physiol 145(4):1647–1657

Cheng SH, Zhuang JY, Fan YY, Du JH, Cao LY (2007a) Progress in research and development on hybrid rice: a super-domesticate in China. Ann Bot 100(5):959–966

Chettry U, Chrungoo NK, Kulkarni K (2019) Comparative transcriptomics approach in elucidation of carotenoid biosynthesis regulation in grains of rice (*Oryza sativa* L.). Sci Rep 9(1):1–12

Chrastil J, Zarins ZM (1992) Influence of storage on peptide subunit composition of rice Oryzenin. J Agric Food Chem 40:927–930

Cicero AF, Gaddi A (2001) Rice bran oil and γ-oryzanol in the treatment of hyperlipoproteinaemias and other conditions. Phytother Res 15(4):277–289

Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc B Biol Sci 363(1491):557–572

Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, De Los Campos G, Burgueño J, González-Camacho JM, Pérez-Elizalde S, Beyene Y, Dreisigacker S (2017) Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci 22 (11):961–975

Crowell S, Korniliev P, Falcao A, Ismail A, Gregorio G, Mezey J, McCouch S (2016) Genome-wide association and high-resolution phenotyping link *Oryza sativa* panicle traits to numerous trait-specific QTL clusters. Nat Commun 7(1):1–14

Dalrymple DG (1986) Development and spread of high-yielding rice varieties in developing countries. Int Rice Res Inst

Das GK, Qudhia P (2001) Rice as medicinal plant in Chhattisgarh (India): a survey. Agric Sci Digest 21:204–205

De Datta SK (1981) Principles and practices of rice production. Int Rice Res Inst, Philipines

Demorest ZL, Coffman A, Baltes NJ et al (2016) Direct stacking of sequence-specific nuclease-induced mutations to produce high oleic and low linolenic soybean oil. BMC Plant Biol 16:225. https://doi.org/10.1186/s12870-016-0906-1

Descalsota GIL, Swamy BM, Zaw H, Inabangan-Asilo MA, Amparado A, Mauleon R, Chadha-Mohanty P, Arocena EC, Raghavan C, Leung H, Hernandez JE (2018) Genome-wide association mapping in a rice MAGIC plus population detects QTLs and genes useful for biofortification. Front Plant Sci 9:1347

Descalsota-Empleo GI, Amparado A, Inabangan-Asilo MA, Tesoro F, Stangoulis J, Reinke R, Swamy BM (2019) Genetic mapping of QTL for agronomic traits and grain mineral elements in rice. Crop J 7(4):560–572

Díaz-Benito P, Banakar R, Rodríguez-Menéndez S, Capell T, Pereiro R, Christou P, Abadía J, Fernández B, Álvarez-Fernández A (2018) Iron and zinc in the embryo and endosperm of rice (*Oryza sativa* L.) seeds in contrasting 2′-deoxymugineic acid/nicotianamine scenarios. Front. Plant Sci 9:1190

Dimitrovski T, GJorgovska N, Levkov V et al (2021) Physicochemical grain properties of the weedy rice complex in North Macedonia. Bulgarian J Agric Sci 27(4):798–803

Dixit S, Singh UM, Abbai R, Ram T, Singh VK, Paul A, Virk PS, Kumar A (2019) Identification of genomic region (s) responsible for high iron and zinc content in rice. Sci Rep 9(1):1–8

Dong W, Thomas N, Ronald PC, Goyer A (2016) Overexpression of thiamin biosynthesis genes in Rice increases leaf and unpolished grain thiamin content but not resistance to *Xanthomonas oryzae pv. oryzae*. Front Plant Sci 7:616

Du J, Zeng D, Wang B, Qian Q, Zheng S, Ling HQ (2013) Environmental effects on mineral accumulation in rice grains and identification of ecological specific QTLs. Environ Geochem Health 35(2):161–170

Eggum BO (1969) Evaluation of protein quality and the development of screening technique. In: New approaches to breeding for improved plant protein. IAEA, Vienna, pp 125–135

Eggum BO (1977) Nutritional aspects of cereal protein. Genetic diversity in plants. Plenum Press, York, pp 349–369

Eggum BO (1979) The nutritional value of rice in comparison with other cereals. In: Proceedings of workshop on chemical aspects of rice grain quality. IRRI, Los Bunos, pp 91–111

Englyst HN, Cummings JH (1985) Digestion of the polysaccharides of some cereal foods in the human small intestine. Am J Clin Nutr 42:778–787

Englyst HN, Kingman SM, Cummings JH (1992) Classification and measurement of nutritionally important starch fractions. Eur J Clin Nutr 46(Suppl 2):S33–S50

Fan F, Long W, Liu M, Yuan H, Pan G, Li N, Li S (2019) Quantitative trait locus mapping of the combining ability for yield-related traits in wild rice *Oryza longistaminata*. J Agric Food Chem 67(32):8766–8772

FAO (2021) The state of food and agriculture. Making agri food systems more resilient to shocks and stresses. FAO, Rome. https://doi.org/10.4060/cb4476en

Fitzgerald MA, Rahman S, Resurreccion AP, Concepcion J, Daygon VD, Dipti SS, Kabir KA, Klingner B, Morell MK, Bird AR (2011) Identification of a major genetic determinant of glycaemic index in rice. Rice 4(2):66–74

Frank T, Meuleye BS, Miller A et al (2007) Metabolite profiling of two low phytic acid (lpa) rice mutants. J Agric Food Chem 55:11011–11019. https://doi.org/10.1021/jf0723559

Frank T, Habernegg R, Yuan F-J et al (2009) Assessment of the contents of phytic acid and divalent cations in low phytic acid (lpa) mutants of rice and soybean. J Food Compos Anal 22:278–284. https://doi.org/10.1016/j.jfca.2008.11.022

Frei M, Becker K (2004) On rice, biodiversity & nutrients. Univ of Hohenheim, Stuttgart. http://www.greenpeaceweb.org/gmo/nutrients.pdf

Fukagawa NK, Ziska LH (2019) Rice: importance for global nutrition. J Nutr Sci Vitaminol 65: S2–S3

Garcia-Oliveira AL, Tan L, Fu Y, Sun C (2009) Genetic identification of quantitative trait loci for contents of mineral nutrients in rice grain. J Integr Plant Biol 51(1):84–92

Garg M, Sharma A, Vats S, Tiwari V, Kumari A, Mishra V, Krishania M (2021) Vitamins in cereals: a critical review of content, health effects, processing losses, bioaccessibility, fortification, and biofortification strategies for their improvement. Front Nutr 8:254

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169(3):1631–1638

Ghasemzadeh A, Karbalaii MT, Jaafar HZ, Rahmat A (2018) Phytochemical constituents, antioxidant activity, and antiproliferative properties of black, red, and brown rice bran. Chem Cent J 12 (1):1–13

Goto F, Yoshihara T, Shigemoto N, Toki S, Takaiwa F (1999) Iron fortification of rice seed by the soybean ferritin gene. Nat Biotechnol 17(3):282–286

Goufo P, Trindade H (2014) Rice antioxidants: phenolic acids, flavonoids, anthocyanins, proanthocyanidins, tocopherols, tocotrienols, γ-oryzanol, and phytic acid. Food Sci Nutr 2:75–104. https://doi.org/10.1002/FSN3.86

GRiSP (2013) Rice almanac, 4th edn. GRiSP (Global Rice Science Partnership), International Rice Research Institute, Los Baños

Gross BL, Reagon M, HSU SC, Caicedo AL, Jia Y, Olsen KM (2011) Seeing red: the origin of grain pigmentation in US weedy rice. Mol Ecol 19(16):3380–3393

Gunaratne A, Wu K, Li D, Bentota A, Corke H, Cai Y (2013) Antioxidant activity and nutritional quality of traditional red-grained rice varieties containing proanthocyanidins. Food Chem 138: 1153–1161. https://doi.org/10.1016/j.foodchem.2012.11.129

Guzman MKD, Parween S, Butardo VM et al (2017) Investigating glycemic potential of rice by unraveling compositional variations in mature grain and starch mobilization patterns during seed germination. Sci Rep 7:5854. https://doi.org/10.1038/s41598-017-06026-0

Gyani PC, Bollinedi H, Krishnan SG et al (2021) Characterizing the diversity of grain phytic acid content in Indian rice landraces by multivariate genetic analyses. Indian. J Genet 81:34–42. https://doi.org/10.31742/IJGPB.81.1.3

Hanamaratti NG, Sathisha TN, Veeresha BA, Surendra P, Salimath PM (2015) Stability analysis of protein content in traditional land races of rice year. Ecoscan 9(1&2):481–485

Hao L, Ruiying Q, Xiaoshuang L, Shengxiang L, Rongfang X, Jianbo Y et al (2019) CRISPR/Cas9-mediated adenine base editing in rice genome. Rice Sci 26(2):125–128. https://doi.org/10.1016/j.rsci.2018.07.002

Hasan MM, Rafii MY, Ismail MR et al (2015) Marker-assisted backcrossing: a useful method for rice improvement. Biotechnol Biotechnol Equip 29:237

Hedden P (2003) Constructing dwarf rice. Nat Biotechnol 21(8):873–874

Hemamalini S, Umamaheswari DS, Lavanya DR, Umamaheswara RDC (2018) Exploring the therapeutic potential and nutritional properties of 'Karuppu Kavuni' variety rice of Tamil Nadu. Int J Pharma Biol Sci 9:88–96

Hiramitsu T, Armstrong D (1991) Preventive effect of antioxidants on lipid peroxidation in the retina. Ophthalmic Res 23:196–203

Hou Z, Qin P, Zhang Y et al (2013) Identification of anthocyanins isolated from black rice (Oryza sativa L.) and their degradation kinetics. Food Res Int 50:691–697

Hu ZL, Li P, Zhou MQ, Zhang ZH, Wang LX, Zhu LH et al (2004) Mapping of quantitative trait loci (QTLs) for rice protein and fat content using doubled haploid lines. Euphytica 135:47

Hu BL, Huang DR, Xiao YQ, Fan YY, Chen DZ, Zhuang JY (2016) Mapping QTLs for mineral element contents in brown and milled rice using an *Oryza sativa* × *O. rufipogon* backcross inbred line population. Cereal Res Commun 44:57–68. https://doi.org/10.1556/0806.43.2015.044

Huang X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42(11): 961–967

Huang X, Zhao Y, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C, Fan D, Lu Y (2012) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat Genet 44(1):32–39

Huang X, Yang S, Gong J, Zhao Y, Feng Q, Gong H, Li W, Zhan Q, Cheng B, Xia J, Chen N (2015) Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. Nat Commun 6(1):1–9

Huang L, Li Q, Zhang C, Chu R, Gu Z, Tan H et al (2020) Creating novel Wx alleles with fine-tuned amylose levels and improved grain quality in rice by promoter editing using CRISPR/Cas9 system. Plant Biotechnol J 18(11):2164–2166

Hui S, Li H, Mawia AM, Zhou L, Cai J, Ahmad S, Lai C, Wang J, Jiao G, Xie L, Shao G (2022) Production of aromatic three-line hybrid rice using novel alleles of BADH2. Plant Biotechnol J 20(1):59–74

Hwang JE, Ahn J-W, Kwon S-J et al (2014) Selection and molecular characterization of a high tocopherol accumulation rice mutant line induced by gamma irradiation. Mol Biol Rep 41: 7671–7681. https://doi.org/10.1007/s11033-014-3660-1

Indurkar AB, Majgahe SK, Sahu VK, Vishwakarma A, Premi V, Shrivastatva P (2015) Identification characterization and mapping of QTLs related to grain Fe, Zn and protein contents in Rice (*Oryza sativa* L.). Elect J Plant Breed 6:1059

Insuan O, Chariyakornkul A, Rungrote Y, Wongpoomchai R (2017) Antimutagenic and antioxidant activities of thai rice brans. J Cancer Prev 22:89–97. https://doi.org/10.15430/JCP.2017.22.2.89

International Rice Genome Sequencing Project (IRGSP) (2005) The map-based sequence of the rice genome. Nature 436:793–800

Irakli MN, Samanidou VF, Katsantonis DN, Billiaderis CG, Papadoyannis IN (2016) Phytochemical profiles and antioxidant capacity of pigmented and non-pigmented genotypes of rice (*Oryza sativa* L.). Cereal Res Commun 44:98–110. https://doi.org/10.1556/0806.43.2015.033

Isaac RSR, Nair AS, Varghese E, Chavali M (2012) Phytochemical, antioxidant and nutrient analysis of medicinal rice (*Oryza sativa* L.) varieties found in South India. Adv Sci Lett 11: 86–90. https://doi.org/10.1166/asl.2012.2174

Ishikawa R, Iwata M, Taniko K, Monden G, Miyazaki N, Orn C et al (2017) Detection of quantitative trait loci controlling grain zinc concentration using Australian wild rice (*Oryza meridionalis*) a potential genetic resource for biofortification of rice. PLoS One 12:e0187224

Ishimaru Y, Masuda H, Bashir K et al (2010) Rice metal-nicotianamine transporter, OsYSL2, is required for the long-distance transport of iron and manganese. Plant J 62:379–390. https://doi.org/10.1111/j.1365-313X.2010.04158.x

Islam W, Qasim M, Noman A, Idrees A, Wang L (2017) Genetic resistance in chickpea against Ascochyta blight: historical efforts and recent accomplishments. J Anim Plant Sci 27(6): 1941–1957

Jacquemin J, Bhatia D, Singh K, Wing RA (2013) The International *Oryza* Map Alignment Project: development of a genus-wide comparative genomics platform to help solve the 9 billion-people question. Curr Opin Plant Biol 16(2):147–156

Jagadeesh BR, Krishnamurthy R, Surekha K, Yogesh GS (2013) Studies on high accumulation of iron and zinc contents in some selected rice genotypes. Global J Bio Biotechnol 2:539–541

Jang S, Han JH, Lee YK, Shin NH, Kang YJ, Kim CK, Chin JH (2020) Mapping and validation of QTLs for the amino acid and total protein content in brown rice. Front Genet 11:240

Jenkins DJ, Wolever TM, Taylor RH, Barker H, Fielden H, Baldwin JM, Bowling AC, Newman HC, Jenkins AL, Goff DV (1981) Glycemic index of foods: a physiological basis for carbohydrate exchange. Am J Clin Nutr 34(3):362–366

Jesch ED, Carr TP (2017) Food ingredients that inhibit cholesterol absorption. Prev Nutr Food Sci 22:67–80. https://doi.org/10.3746/pnf.2017.22.2.67

Jia L, Xie L, Lao S, Zhu QH, Fan L (2021) Rice bioinformatics in the genomic era: status and perspectives. Crop J 9(3):609–621

Jiang C, Cheng Z, Zhang C et al (2014) Proteomic analysis of seed storage proteins in wild rice species of the *Oryza* genus. Proteome Sci 12:51

Jiang S-Y, Ma A, Xie L, Ramachandran S (2016) Improving protein content and quality by oer-expressing artificially synthetic fusion proteins with high lysine and threonine constituent in rice plants. Sci Rep 6:34427

Johnson AA, Kyriacou B, Callahan DL, Carruthers L, Stangoulis J, Lombi E, Tester M (2011) Constitutive overexpression of the OsNAS gene family reveals single-gene strategies for effective iron-and zinc-biofortification of rice endosperm. PLoS One 6(9):e24476

Jukanti AK, Pautong PA, Liu Q, Sreenivasulu N (2020) Low glycemic index rice – a desired trait in starchy staples. Trends Food Sci Technol 106:132–149

Juliano BO (1971) A simplified assay for milled rice amylose. Cereal Sci Today 16:334–360
Juliano BO (1977) Rice lipids. Riso 26:3–21
Kalaivani R, Arulmozhi P, Bakiyalakshmi SV (2018) A study on medicinal properties of traditional rice karung kavuni and neutraceutical formulation. Int J Food Nutri Sci 5(1):86–90
Kalita T, Hazarika J (2022) Morphological study, nutritional value and phytochemical estimation of four indigenous rice (*Oryza sativa* L.) varieties of Assam, India. Food Res 6:196–203
Kamiloglu S, Capanoglu E, Grootaert C, van Camp J (2015) Anthocyanin absorption and metabolism by human intestinal Caco-2 cells – a review. Int J Mol Sci 16:21555. https://doi.org/10.3390/IJMS160921555
Kaneko R, Tsuchiya T (1955) New compound in rice bran and germ oils. Chem Abst 49:4897
Kang K, Jang S-M, Kang S, Back K (2005) Enhanced neutraceutical serotonin derivatives of rice seed by hydroxycinnamoyl-CoA:serotonin N-(hydroxycinnamoyl) transferase. Plant Sci 168:783–788
Kappara S, Neelamraju S, Ramanan R (2018) Down regulation of a heavy metal transporter gene influences several domestication traits and grain Fe-Zn content in rice. Plant Sci 276:208–219
Kar S, Mai HJ, Khalouf H, Ben Abdallah H, Flachbart S, Fink-Straube C, Bräutigam A, Xiong G, Shang L, Panda SK, Bauer P (2021) Comparative transcriptomics of lowland rice varieties uncovers novel candidate genes for adaptive Fe excess tolerance. Plant Cell Physiol 62(4):624–640
Kawahigashi H, Hirose S, Iwai T et al (2007) Chemically induced expression of rice OSB2 under the control of the OsPR1.1 promoter confers increased anthocyanin accumulation in transgenic rice. J Agric Food Chem 55:1241–1247
Kawakami Y, Gruissem W, Bhullar NK (2022) Novel rice iron biofortification approaches using expression of ZmYS1 and OsTOM1 controlled by tissue-specific promoters. J Exp Bot 73(16):5440–5459. https://doi.org/10.1093/jxb/erac214
Kawakatsu T, Wang S, Wakasa Y, Takaiwa F (2010) Increased lysine content in rice grains by over-accumulation of BiP in the endosperm. Biosci Biotechnol Biochem 74:2529–2531. https://doi.org/10.1271/bbb.100619
Khan MSS, Basnet R, Islam SA, Shu Q (2019) Mutational analysis of OsPLDα1 reveals its involvement in phytic acid biosynthesis in rice grains. J Agric Food Chem 67(41):11436–11443. https://doi.org/10.1021/acs.jafc.9b05052
Khatoon M, Islam MT (2020) Grain shape, protein, zinc and iron content of rice land races in Bangladesh. Int J Expt Agric 10(2):7–11
Khush GS (1987) Rice breeding: past, present and future. J Genet 66(3):195–216
Khush GS (1997) Origin, dispersal, cultivation and variation of rice. Plant Mol Biol 35(1):25–34
Kim SI, Andaya CB, Newman JW et al (2008) Isolation and characterization of a low phytic acid rice mutant reveals a mutation in the rice orthologue of maize MIK. Theor Appl Genet 117:1291–1301
Kim GR, Jung ES, Lee S, Lim S, Ha S, Lee CH (2014) Combined mass spectrometry-based metabolite profiling of different pigmented rice (*Oryza sativa* L.) seeds and correlation with antioxidant activities. Molecules 19:15673–15686. https://doi.org/10.3390/molecules191015673
Kline K, Yu W, Sanders BG (2004) Vitamin E and breast cancer. J Nutr 134(12):3458S–3462S
Kobayashi T, Nagasaka S, Senoura T, Itai RN, Nakanishi H, Nishizawa NK (2013) Iron-binding haemerythrin RING ubiquitin ligases regulate plant iron responses and accumulation. Nat Commun 4(1):1–2
Kobayashi T, Ozu A, Kobayashi S, An G, Jeon JS, Nishizawa NK (2019) OsbHLH058 and OsbHLH059 transcription factors positively regulate iron deficiency responses in rice. Plant Mol Biol 101(4):471–486
Kobayashi T, Nagano AJ, Nishizawa NK (2021) Iron deficiency-inducible peptide-coding genes OsIMA1 and OsIMA2 positively regulate a major pathway of iron uptake and translocation in rice. J Exp Bot 72(6):2196–2211
Kole C, Muthamilarasan M, Henry R, Edwards D, Sharma R, Abberton M, Batley J, Bentley A, Blakeney M, Bryant J, Cai H et al (2015) Application of genomics-assisted breeding for generation of climate resilient crops: progress and prospects. Front Plant Sci 6:563

Kroon PA, Clifford MN, Crozier A, Day AJ, Donovan JL, Manach C, Williamson G (2004) How should we assess the effects of exposure to dietary polyphenols in vitro? Am J Clin Nutr 80(1): 15–21

Kumar Ashok S, Jaganathan D, Ramanathan V, Rahman H, Palaniswamy R, Kambale R, Muthurajan R (2020) Creation of novel alleles of fragrance gene *OsBADH2* in rice through CRISPR/Cas9 mediated gene editing. PLoS One 15(8):e0237018

Kumar J, Jain S, Jain RK (2014) Linkage mapping for grain iron and zinc content in F2 population derived from the cross between PAU 201 and Palman 579 in rice (*Oryza sativa* L.). Cereal Res Commun 42:389

Kumar N, Jain RK, Chowdhury VK (2019) Linkage mapping of QTLs for grain minerals (iron and zinc) and physio-morphological traits for development of mineral rich rice (*Oryza sativa* L.). Indian J Biotechnol 18:69–80

Kumar D, Yadav A, Ahmad R, Dwivedi UN, Yadav K (2022) CRISPR-based genome editing for nutrient enrichment in crops: a promising approach toward global food security. Front Genet 13:1650

Kusmiyati F, Lukiwati DR, Kristanto BA, Herwibawa B (2019) Glycemic index of ten commercially Indonesian rice cultivars. IOP Conf Ser: Earth Environ Sci 250:012028. https://doi.org/10.1088/1755-1315/250/1/012028

Kuwano M, Mimura T, Takaiwa F, Yoshida KT (2009) Generation of stable 'low phytic acid' transgenic rice through antisense repression of the 1d-myo-inositol 3-phosphate synthase gene (RINO1) using the 18-kDa oleosin promoter. Plant Biotechnol J 7:96–105

Lam PY, Liu H, Lo C (2015) Completion of tricin biosynthesis pathway in rice: cytochrome P450 75B4 is a unique chrysoeriol 5′-hydroxylase. Plant Physiol 168:1527–1536

Lang NT, Nguyet TA, Phang NV, Buu BC (2007) Breeding for low phytic acid mutants in rice (*Oryza sativa L.*). Omonrice 15:29–35

Lap B, Rai M, Tyagi W (2021) Playing with colours: genetics and regulatory mechanisms for anthocyanin pathway in cereals. Biotechnol Genet Eng Rev 37(1):1–29

Lee S, An G (2009) Over-expression of OsIRT1 leads to increased iron and zinc accumulations in rice. Plant Cell Environ 32(4):408–416

Lee S, Chiecko JC, Kim SA, Walker EL, Lee Y, Guerinot ML, An G (2009a) Disruption of OsYSL15 leads to iron inefficiency in rice plants. Plant Physiol 150(2):786–800

Lee S, Jeon US, Lee SJ, Kim YK, Persson DP, Husted S, Schjorring JK, Kakei Y, Masuda H, Nishizawa NK, Ana G (2009b) Iron fortification of rice seeds through activation of the nicotianamine synthase gene. Proc Natl Acad Sci U S A 106:22014–22019

Lee S, Kim YS, Jeon US, Kim YK, Schjoerring JK, An G (2012) Activation of rice nicotianamine synthase 2 (OsNAS2) enhances iron availability for biofortification. Mol Cells 33(3):269–275

Lee GH, Yun BW, Kim KM (2014) Analysis of QTLs associated with the rice quality related gene by double haploid populations. Int J Genom:781832. https://doi.org/10.1155/2014/781832

Lee JS, Sreenivasulu N, Hamilton RS, Kohli A (2019) Brown rice, a diet rich in health promoting properties. J Nutr Sci Vitaminol 65:S26–S28

Lei GJ, Yamaji N, Ma JF (2021) Two metallothionein genes highly expressed in rice nodes are involved in distribution of Zn to the grain. New Phytol 229(2):1007–1020

Leng Y, Xue D, Yang Y, Hu S, Su Y, Huang L et al (2014) Mapping of QTLs for eating and cooking quality-related traits in rice (*Oryza sativa* L.). Euphytica 197:99

Li J, Xiao J, Grandillo S, Jiang L, Wan Y, Deng Q et al (2004) QTL detection for rice grain quality traits using an interspecific backcross population derived from cultivated Asian (*O. sativa* L.) and African (*O. glaberrima* S.) rice. Genome 47:697

Li Y, Xiao J, Chen L, Huang X, Cheng Z, Han B, Zhang Q, Wu C (2018) Rice functional genomics research: past decade and future. Mol Plant 11(3):359–380

Li Q, Chen L, Yang A (2020) The molecular mechanisms underlying Fe deficiency responses in rice. Int J Mol Sci 21(1):43

Lin P-Y, Lai H-M (2011) Bioactive compounds in rice during grain development. Food Chem 127: 86–93. https://doi.org/10.1016/j.foodchem.2010.12.092

Liu JC, Irene O, Michael T, John NAL (2004) Phytic acid-phosphorus and other nutritionally important mineral nutrient elements in grains of wild-type and low phytic acid (lpa1–1) rice. Seed Sci Res 14:109–116

Liu X, Zhang C, Wang X et al (2016) Development of high-lysine rice via endosperm-specific expression of a foreign lysine rich protein gene. BMC Plant Biol 16:147. https://doi.org/10.1186/s12870-016-0837-x

Liu X, Fan F, Liu M, Long W, Yu Y, Yuan H, Pan G, Li N, Li S, Liu J (2020) Quantitative trait loci mapping of mineral element contents in brown rice using backcross inbred lines derived from *O. longistaminata*. Front Plant Sci 11:1229

Lloyd BJ, Siebenmorgen TJ, Beers KW (2000) Effects of commercial processing on antioxidants in rice bran. Cereal Chem 77(5):551–555

Londo JP, Chiang YC, Hung KH, Chiang TY, Schaal BA (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. Proc Natl Acad Sci U S A 103(25):9578–9583

Longvah T, Prasad VSS (2020) Nutritional variability and milling losses of rice landraces from Arunachal Pradesh, Northeast India. Food Chem 318:126385

Longvah T, Babu VR, Viraktamath BC (2012) Nutrient diversity within rice cultivars *(Oryza sativa L)* from India. In: Sustainable diets and biodiversity: directions and solutions for policy, research and action. FAO, Biodiversity International, Rome

Longvah T, Anantan I, Bhaskarachary K, Venkaiah K, Longvah T (2017) Indian food composition tables. Hyderabad, National Institute of Nutrition, Indian Council of Medical Research, pp 2–58

Lou J, Chen L, Yue G, Lou Q, Mei H, Xiong L et al (2009) QTL mapping of grain quality traits in rice. J Cereal Sci 50:145

Lu K, Li L, Zheng X, Zhang Z, Mou T, Hu Z (2008) Quantitative trait loci controlling Cu, Ca, Zn, Mn and Fe content in rice grains. J Genet 87:305

Lu K, Li L, Zheng X, Zhang Z, Mou T, Hu Z (2009) Genetic dissection of amino acid content in rice grain. J Food Sci Agric 89:2377

Lucca P, Hurrell R, Potrykus I (2001) Genetic engineering approaches to improve the bioavailability and the level of iron in rice grains. Theor Appl Genet 102(2):392–397

Ma X, Zhang Q, Zhu Q et al (2015) A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. Mol Plant 8:1274–1284. https://doi.org/10.1016/J.MOLP.2015.04.007

Mackon E, Jeazet Dongho Epse Mackon GC, Ma Y et al (2021) Recent insights into anthocyanin pigmentation, synthesis, trafficking, and regulatory mechanisms in rice (*Oryza sativa* L.) caryopsis. Biomol Ther 11:1–26

Mahender A, Anandan A, Pradhan SK, Pandit E (2016) Rice grain nutritional traits and their enhancement using relevant genes and QTLs through advanced approaches. Springerplus 5 (1):1–18

Mangel N, Fudge JB, Li K-T et al (2019) Enhancement of vitamin B6 levels in rice expressing Arabidopsis vitamin B6 biosynthesis de novo genes. Plant J 99:1047–1065. https://doi.org/10.1111/tpj.14379

Mansueto L, Fuentes RR, Borja FN, Detras J, Abriol-Santos JM, Chebotarov D, Sanciangco M, Palis K, Copetti D, Poliakov A, Dubchak I (2017) Rice SNP-seek database update: new SNPs, indels, and queries. Nucleic Acids Res 45(D1):D1075–D1081

Marles RJ (2017) Mineral nutrient composition of vegetables, fruits and grains: the context of reports of apparent historical declines. J Food Compos Anal 56:93–103

Masuda H, Usuda K, Kobayashi T, Ishimaru Y, Kakei Y, Takahashi M, Higuchi K, Nakanishi H, Mori S, Nishizawa NK (2009) Overexpression of the barley nicotianamine synthase gene HvNAS1 increases iron and zinc concentrations in rice grains. Rice 2(4):155–166

Masuda H, Kobayashi T, Ishimaru Y, Takahashi M, Aung MS, Nakanishi H, Mori S, Nishizawa NK (2013) Iron-biofortification in rice by the introduction of three barley genes participated in mugineic acid biosynthesis with soybean ferritin gene. Front Plant Sci 4:132

Mbanjo EGN, Kretzschmar T, Jones H et al (2020) The genetic basis and nutritional benefits of pigmented rice grain. Front Genet 11:229. https://doi.org/10.3389/fgene.2020.00229
McCouch SR, Wright MH, Tung CW, Maron LG, McNally KL, Fitzgerald M, Singh N, DeClerck G, Agosto-Perez F, Korniliev P, Greenberg AJ (2016) Open access resources for genome-wide association mapping in rice. Nat Commun 7(1):1–4
McDonald GK, Genc Y, Graham RD (2008) A simple method to evaluate genetic variation in grain Zn concentration by correcting for differences in grain yield. Plant Soil 306:49
Melini V, Acquistucci R (2017) Health-promoting compounds in pigmented Thai and wild rice. Foods 6(1):9
Meuwissen TH, Hayes BJ, Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4):1819–1829
Miura S, Koyama N, Crofts N et al (2021) Generation and starch characterization of non-transgenic bei and beiib double mutant rice (*Oryza sativa*) with ultra-high level of resistant starch. Rice 14: 3. https://doi.org/10.1186/s12284-020-00441-0
Mondal D, Kantamraju P, Jha S et al (2021) Evaluation of indigenous aromatic rice cultivars from sub-Himalayan Terai region of India for nutritional attributes and blast resistance. Sci Rep 11: 4786. https://doi.org/10.1038/s41598-021-83921-7
Mussurova S, Al-Bader N, Zuccolo A, Wing RA (2020) Potential of platinum standard reference genomes to exploit natural variation in the wild relatives of rice. Front Plant Sci 11:579980. https://doi.org/10.3389/fpls.2020.579980
Muthayya S, Sugimoto JD, Montgomery S, Maberly GF (2014) An overview of global rice production, supply, trade, and consumption. Ann N Y Acad Sci 1324(1):7–14
Nagamine A, Ezura H (2022) Genome editing for improving crop nutrition. Front Genome Ed 4: 850104. https://doi.org/10.3389/fgeed.2022.850104
Neeraja CN, Babu VR, Ram S, Hossain F, Hariprasanna K, Rajpurohit BS, Prabhakar, Longvah T, Prasad KS, Sandhu JS, Datta SK (2017) Biofortification in cereals: progress and prospects. Curr Sci 25:1050–1057
Neeraja CN, Kulkarni KS, Madhu Babu P, Sanjeeva Rao D, Surekha K, Ravindra Babu V (2018) Transporter genes identified in landraces associated with high Zn in polished rice through panicle transcriptome for biofortification. PLoS One 13(2):e0192362
Neeraja CN, Hossain F, Hariprasanna K, Ram S, Satyavathi CT, Longvah T, Raghu P, Voleti SR, Sundaram RM (2022) Towards nutrition security of India with biofortified cereal varieties. Curr Sci 123(3):271–277
Nguyen HC, Hoefgen R, Hesse H (2012) Improving the nutritive value of rice seeds: elevation of cysteine and methionine contents in rice plants by ectopic expression of a bacterial serine acetyltransferase. J Exp Bot 63:5991–6001. https://doi.org/10.1093/jxb/ers253
Norton GJ, Deacon CM, Xiong L, Huang S, Meharg AA, Price AH (2010) Genetic mapping of the rice genome in leaves and grain: identification of QTLs for 17 elements including arsenic, cadmium, iron and selenium. Plant Soil 329:139
Norton GJ, Douglas A, Lahner B, Yakubova E, Guerinot ML, Pinson SR, Tarpley L, Eizenga GC, McGrath SP, Zhao FJ, Islam MR (2014) Genome wide association mapping of grain arsenic, copper, molybdenum and Zn in rice (*Oryza sativa* L.) grown at four international field sites. PLoS One 9(2):e89685
Nounmusig J, Kongkachuichai R, Sirichakwal PP et al (2018) The effect of low and high glycemic index based rice varieties in test meals on postprandial blood glucose, insulin and incretin hormones response in prediabetic subjects. Int Food Res J 25:835–841
Nozoye T, Nagasaka S, Kobayashi T, Takahashi M, Sato Y, Sato Y, Uozumi N, Nakanishi H, Nishizawa NK (2011) Phytosiderophore efflux transporters are crucial for iron acquisition in graminaceous plants. J Biol Chem 286(7):5446–5454
Nugent AP (2005) Health properties of resistant starch. Nutr Bull 30:27–54. https://doi.org/10.1111/J.1467-3010.2005.00481.X
Ogawa M, Kumamaru T, Satoh H et al (1987) Purification of protein body-I of rice seed and its polypeptide composition. Plant Cell Physiol 28:1517–1527

Ogo Y, Itai RN, Kobayashi T, Sann Aung M, Nakanishi H, Nishizawa NK (2011a) OsIRO2 is responsible for iron utilization in rice and improves growth and yield in calcareous soil. Plant Mol Biol 75(6):593–605

Oliva N, Chadha-Mohanty P, Poletti S, Abrigo E, Atienza G, Torrizo L, Garcia R, Dueñas C, Poncio MA, Balindong J, Manzanilla M (2014) Large-scale production and evaluation of marker-free indica rice IR64 expressing phytoferritin genes. Mol Breed 33(1):23–37

Oliva N, Florida Cueto-Reaño M, Trijatmiko KR, Samia M, Welsch R, Schaub P, Beyer P, Mackenzie D, Boncodin R, Reinke R, Slamet-Loedin I (2020) Molecular characterization and safety assessment of biofortified provitamin A rice. Sci Rep 10(1):1–13

Owens B (2018) Golden Rice is safe to eat, says FDA. Nat Biotechnol 36(7):559

Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, Vernon G, Wright SY, Hinchliffe E, Adams JL, Silverstone AL, Drake R (2005) Improving the nutritional value of Golden Rice through increased pro-vitamin A content. Nat Biotechnol 23(4):482–487

Park S, Choi MJ, Lee JY, Kim JK, Ha SH, Lim SH (2016) Molecular and biochemical analysis of two rice flavonoid 3′-hydroxylase to evaluate their roles in flavonoid biosynthesis in rice grain. Int J Mol Sci 17(9):1549

Parrado J, Miramontes E, Jover M et al (2006) Preparation of a rice bran enzymatic extract with potential use as functional food. Food Chem 98:742–748

Pereira-Caro G, Watanabe S, Crozier A, Fujimura T, Yokota T (2013) Phytochemical profile of a Japanese black – purple rice. Food Chem 141:2821–2827

Petroni K, Landoni M, Tomay F, Calvenzani V, Simonelli C, Cormegna M (2017) Proximate composition, polyphenol content and anti-inflammatory properties of white and pigmented Italian rice varieties. Univ J Agric 5(5):312–321

Pippal A, Bhusal N, Meena RK, Bishnoi M, Bhoyar PI, Jain RK (2022) Identification of genomic locations associated with grain micronutrients (Fe and Zn) in rice (*Oryza sativa* L.). Genet Resour Crop Evol 69(1):221–230

Potrykus I (2001) The 'Golden rice' tale. In Vitro Cell Dev Biol Plant 37:93–100. https://doi.org/10.1007/S11627-001-0019-9

Poulev A, Heckman JR, Raskin I, Belanger FC (2019) Tricin levels and expression of flavonoid biosynthetic genes in developing grains of purple and brown pericarp rice. Peer J 7:e6477. https://doi.org/10.7717/peerj.6477

Pradhan SK, Pandit E, Pawar S, Naveenkumar R, Barik SR, Mohanty SP, Nayak DK, Ghritlahre SK, Sanjiba Rao D, Reddy JN, Patnaik SSC (2020) Linkage disequilibrium mapping for grain Fe and Zn enhancing QTLs useful for nutrient dense rice breeding. BMC Plant Biol 20 (1):1–24

Pramai P, Jiamyangyuen S (2016) Chemometric classification of pigmented rice varieties based on antioxidative properties in relation to color. Songklanakarin J Sci Technol 38:463–472

Prasanthi PS, Jeyakumar SM, Panda H, Srinivas E, Damayanti K (2019) Development of low glycemic index rice and its effectiveness in the regulation of postprandial glucose response in type 2 diabetes. Am J Food Nutr 7:158–165

Priyadarshi R, Sinha P, Dasari A, Sundaram RM (2021) Application of "omics" technologies in crop breeding. In: Kumar A, Kumar R, Shukla P, Patel HK (eds) Omics technologies for sustainable agriculture and global food security, vol II. Springer, Singapore, pp 25–45

Pushpam R, Mythili SR, Nikitha TC (2019) Medicinal rice and its medicinal values. Int J Curr Microbiol Appl Sci 8:2090–2095

Qamar ZU, Hameed A, Ashraf M, Rizwan M, Akhtar M (2019) Development and molecular characterization of low phytate basmati rice through induced mutagenesis, hybridization, back-cross, and marker assisted breeding. Front Plant Sci 10:1525. https://doi.org/10.3389/fpls.2019.01525

Qin Y, Kim SM, Sohn JK (2009) QTL analysis of protein content in double haploid lines of rice. Korean J Crop Sci 54:165

Qiu SQ, Lu Q, Yu HH, Ni XM, Zhang GY, He H, Xie WB, Zhou FS (2018) The development and application of rice whole genome selection breeding platform. Chin Bull Life Sci 30:1120–1128

Qu LQ, Yoshihara T, Ooyama A, Goto F, Takaiwa F (2005) Iron accumulation does not parallel the high expression level of ferritin in transgenic rice seeds. Planta 222:225–233

Raghuvanshi RS, Dutta A, Tewari G, Suri S (2017) Qualitative characteristics of red rice and white rice procured from local market of Uttarakhand: a comparative study. J Rice Res 10(1):49–53

Rajendran M, Chandran KR (2020) Grain dimension, nutrition and nutraceutical properties of black and red varieties of rice in India. Curr Res Nutr Food Sci 8:903–923. https://doi.org/10.12944/CRNFSJ.8.3.20

Rathod R, Suman K, Madhubabu P, Subbarao LV, Eswari KB, Chary DS, Neeraja CN (2021) Identification of genomic regions associated with grain zinc concentration in RILs derived from popular cultivar MTU1010 and BR2655 of rice. Indian J Genet 81(04):44–53

Rathod S, Chitikela G, Bandumula N, Ondrasek G, Ravichandran S, Sundaram RM (2022) Modeling and forecasting of rice prices in India during the COVID-19 lockdown using machine learning approaches. Agronomy 12(9):2133

Ratnayake WS, Jackson DS (2008) Thermal behavior of resistant starches RS 2, RS 3, and RS 4. J Food Sci 73. https://doi.org/10.1111/J.1750-3841.2008.00754.X

Ravindra Babu V (2013) Importance and advantages of rice biofortification with iron and zinc. J SAT Agric Res 11:1–6

Raza Q, Riaz A, Sabar M, Atif RM, Bashir K (2019) Meta-analysis of grain Fe and Zn associated QTLs identified hotspot chromosomal regions and positional candidate genes for breeding biofortified rice. Plant Sci 288:110214

Reboul E (2019) Mechanisms of carotenoid intestinal absorption: where do we stand? Nutrients 11. https://doi.org/10.3390/NU11040838

Reddy KD, Bhotmange MG (2013) Isolation of starch from Rice (*Oryza sativa* L.) and its morphological study using scanning electron microscopy. Int J Agric Food Sci Technol 4: 859–866

Rerkasem B, Jumrus S, Yimyam N, Prom-U-Thai C (2015) Variation of grain nutritional quality among Thai purple rice genotypes grown at two different altitudes. Sci Asia 41(6):377–385

Rogers EJ, Rice SM, Nicolosi RJ et al (1993) Identification and quantitation of γ-oryzanol components and simultaneous assessment of tocols in rice bran oil. J Am Oil Chem Soc 70: 301–307. https://doi.org/10.1007/BF02545312

Roy P, Deb D, Pradeep T et al (2021) Comparative analyses of the nutraceutical potentialities of selected Indian traditional black rice (*Oryza sativa* L.) landraces. Oryza 58:295–309

Rutger JN, Raboy V, Moldenhauer K et al (2004) Registration of KBNT lpa1-1 low phytic acid germplasm of rice. Crop Sci 44:363–364

Sahu PK, Mondal S, Sao R, Vishwakarma G, Kumar V, Das BK, Sharma D (2020) Genome-wide association mapping revealed numerous novel genomic loci for grain nutritional and yield-related traits in rice (*Oryza sativa* L.) landraces. 3 Biotech 10(11):1–22

Saika H, Oikawa A, Matsuda F et al (2011) Application of gene targeting to designed mutation breeding of high-tryptophan rice. Plant Physiol 156:1269–1277

Sakulsingharoj C, Inta P, Sukkasem R et al (2014) Overexpression of OSB2 gene in transgenic rice up-regulated expression of structural genes in anthocyanin biosynthesis pathway. Genom Genet 7:173–182. https://doi.org/10.14456/tjg.2014.24

Sanghamitra P, Bagchi TB, Sah RP et al (2017) Characterization of red and purple-pericarp rice (*Oryza sativa* L.) based on physico-chemical and antioxidative properties of grains. Oryza 54 (1):57–64

Sanghamitra P, Sah RP, Bagchi TB et al (2018) Evaluation of variability and environmental stability of grain quality and agronomic parameters of pigmented rice (*Oryza. sativa* L.). J Food Sci Technol 55:879–890. https://doi.org/10.1007/s13197-017-2978-9

Sanjeeva Rao DS, Babu PM, Swarnalatha P et al (2014) Assessment of grain zinc and iron variability in rice germplasm using energy dispersive X-ray fluorescence spectrophotometer (ED – XRF). J Rice Res 7(1&2):45–52

Sanjeeva Rao D, Neeraja CN, Madhu Babu P et al (2020) Zinc biofortified rice varieties: challenges, possibilities, and progress in India. Front Nutr 7:1–13

Sanjeeva Rao DS, Ananthan R, Longvah T et al (2022) Low GI rice for rice consumers suffering from diabetes mellitus. Curr Sci 122:3. https://doi.org/10.3389/fspor.2021.638139

Santos KFDN, Silveira RDD, Martin-Didonet CCG, Brondani C (2013) Storage protein profile and amino acid content in wild rice oryza glumaepatula. Pesq Agropec Bras 48:66–72. https://doi.org/10.1590/S0100-204X2013000100009

Sarkar S, Pattanaik SSC, Chattopadhyay K et al (2018) Genetic improvement of rice for aroma, nutrition and grain quality. In: Pathak H, Nayak AK, Jena M, Singh ON, Samal P, Sharma SG (eds) Rice research for enhancing productivity, profitability and climate resilience. ICAR-NRRI, Cuttack

Schaub P, Wüst F, Koschmieder J et al (2017) Nonenzymatic β-carotene degradation in provitamin A-biofortified crop plants. J Agric Food Chem 65:6588–6598

Schmidhuber J, Sur P, Fay K, Huntley B, Salama J, Lee A, Cornaby L, Horino M, Murray C, Afshin A (2018) The global nutrient database: availability of macronutrients and micronutrients in 195 countries from 1980 to 2013. Lancet Planetary Health 2(8):e353–e368

Selvaraj R, Singh AK, Singh VK et al (2021) Superior haplotypes towards development of low glycemic index rice with preferred grain and cooking quality. Sci Rep 11:10082. https://doi.org/10.1038/s41598-021-87964-8

Senoura T, Sakashita E, Kobayashi T, Takahashi M, Aung MS, Masuda H, Nakanishi H, Nishizawa NK (2017) The iron-chelate transporter OsYSL9 plays a role in iron distribution in developing rice grains. Plant Mol Biol 95(4):375–387

Shao Y, Hu Z, Yu Y, Mou R, Zhu Z, Beta T (2018) Phenolic acids, anthocyanis, proanthocyanidins, antioxidant activity, minerals and their correlations in non-pigmented, red, and black rice. Food Chem 239:733–741

Shen L, Wang C, Fu Y, Wang J, Liu Q, Zhang X et al (2018) QTL editing confers opposing yield performance in different rice varieties. J Integr Plant Biol 60(2):89–93. https://doi.org/10.1111/jipb.12501

Shi CH, Shi Y, Lou XY, Xu HM, Zheng X, Wu JG (2009) Identification of endosperm and maternal plant QTLs for protein and lysine contents of rice across different environments. Crop Pasture Sci 60:295

Shin Y-M, Park H-J, Yim S-D et al (2006) Transgenic rice lines expressing maize C1 and R-S regulatory genes produce various flavonoids in the endosperm. Plant Biotechnol J 4:303–315. https://doi.org/10.1111/j.1467-7652.2006.00182.x

Shozib H (2018) Nutraceutical properties of Bangladeshi rice varieties. Vitam Miner 7:4. https://doi.org/10.4172/2376-1318.1000175

Shozib HB, Bhowmick S, Jahan S et al (2017) In vivo screening for low glycemic index (GI) rice varieties in Bangladesh and evaluate the effect of differently processed rice and rice products on GI. Biol J Sci Technol 5:1–8

Singh N, Jayaswal PK, Panda K, Mandal P, Kumar V, Singh B, Mishra S, Singh Y, Singh R, Rai V, Gupta A (2015) Single-copy gene based 50 K SNP chip for genetic studies and molecular breeding in rice. Sci Rep 5(1):1–9

Sinha P, Singh VK, Bohra A, Kumar A, Reif JC, Varshney RK (2021) Genomics and breeding innovations for enhancing genetic gain for climate resilience and nutrition traits. Theor Appl Genet 134(6):1829–1843

Smith MR, Myers SS (2018) Impact of anthropogenic CO2 emissions on global human nutrition. Nat Clim Chang 8(9):834–839

Smith MR, Myers SS (2019) Global health implications of nutrient changes in rice under high atmospheric carbon dioxide. GeoHealth 3(7):190–200

Soares JC, Santos CS, Carvalho S, Pintado MM, Vasconcelos MW (2019) Preserving the nutritional quality of crop plants under a changing climate: importance and strategies. Plant Soil 443(1):1–26

Soba D, BenMariem S, Fuertes-Mendizábal T, Méndez-Espinoza AM, Gilard F, González-Murua C, Irigoyen JJ, Tcherkez G, Aranjuelo I (2019) Metabolic effects of elevated CO2 on wheat grain development and composition. J Agric Food Chem 67(31):8441–8451

Sookwong P, Nakagawa K, Murata K et al (2007) Quantitation of tocotrienol and tocopherol in various rice brans. J Agric Food Chem 55:461–466. https://doi.org/10.1021/jf0621572

Souci SW, Fuchmann W, Kraut H (1986) Colorimetry of total phenolics with phosphomolybdic-phosphotungstic acid reagents. Food composition and nutrition tables. Wissenschafliche Verlagsgesellschaft, Stuttgart

Sridevi P, Veni BK, Raja DS, Jyothula DPB (2021) Physico-chemical, nutritional and anti-oxidative properties of different colored grain genotypes of rice (*Oryza sativa L.*). Int J Chem Stud 9: 1769–1776

Stangoulis JC, Huynh BL, Welch RM, Choi EY, Graham RD (2007) Quantitative trait loci for phytate in rice grain and their relationship with grain micronutrient content. Euphytica 154(3): 289–294

Sudtasarn G, Homsombat W, Chotechuen S, Chamarerk V (2019) Quantification of tocopherols, tocotrienols and γ-oryzanol contents of local rice varieties in north eastern Thailand. J Nutr Sci Vitaminol 65:S125–S128. https://doi.org/10.3177/jnsv.65.S125

Suman K, Neeraja CN, Madhubabu P, Rathod S, Bej S, Jadhav KP, Kumar JA et al (2021) Identification of promising RILs for high grain zinc through genotype× environment analysis and stable grain zinc QTL using SSRs and SNPs in rice (*Oryza sativa* L.). Front Plant Sci 12: 587482

Sun Y, Jiao G, Liu Z et al (2017) Generation of high-amylose rice through crispr/cas9-mediated targeted mutagenesis of starch branching enzymes. Front Plant Sci 8:298

Sun L, Tan KWJ, Lim JZ et al (2018) Dietary fat and carbohydrate quality have independent effects on postprandial glucose and lipid responses. Eur J Nutr 57:243–250

Sundaram RM, Madhav MS, Neeraja CN, Balachandran SM, Mangrauthia SK, Barbadikar KB, Divya PS, Subba Rao LV, Hariprasad AS (2018) Multi-trait improvement in rice through marker-assisted breeding. Oryza 55:24–37

Sundararajan S, Rajendran V, Sivakumar HP et al (2021) Enhanced vitamin E content in an Indica rice cultivar harbouring two transgenes from *Arabidopsis thaliana* involved in tocopherol biosynthesis pathway. Plant Biotechnol J 19:1083–1085

Suwannakul N, Punvittayagul C, Jarukamjorn K, Wongpoomchai R (2015) Purple rice bran extract attenuates the aflatoxin B1-induced initiation stage of hepatocarcinogenesis by alteration of xenobiotic metabolizing enzymes. Asian Pac J Cancer 16(8):3371–3376

Swamy BM, Descalsota GIL, Nha CT, Amparado A, Inabangan-Asilo MA, Manito C et al (2018a) Identification of genomic regions associated with agronomic and biofortification traits in DH populations of rice. PLoS One 13:e0201756

Swamy BPM, Kaladhar K, Anuradha K, Batchu AK, Longvah T, Sarla N (2018b) QTL analysis for grain Fe and Zn concentrations in two *O. nivara* derived backcross populations. Rice Sci 25(4): 197–207

Takahashi S, Ohtani T, Satoh H et al (2010) Development of coenzyme q10-enriched rice using sugary and shrunken mutants. Biosci Biotechnol Biochem 74:182–184. https://doi.org/10.1271/bbb.90562

Takeuchi Y, Nonoue Y, Ebitani T, Suzuki K, Aoki N, Sato H et al (2007) QTL detection for eating quality including glossiness stickiness taste and hardness of cooked rice. Breed Sci 57:231

Takita T, Higashi T, Yamaguchi M et al (2001) Breeding of a new purple grain rice cultivar "Okuno-murasaki". Bull Tohoku Natl Agric Exp Stn 98:1–10

Tan YF, Sun M, Xing YZ, Hua JP, Sun XL, Zhang QF et al (2001) Mapping quantitative trait loci for milling quality protein content and color characteristics of rice using a recombinant inbred line population derived from an elite rice hybrid. Theor Appl Genet 103(6):1037–1045

Tan S, Han R, Li P, Yang G, Li S, Zhang P, Wang WB, Zhao WZ, Yin LP (2015) Over-expression of the MxIRT1 gene increases iron and zinc content in rice seeds. Transgenic Res 24(1):109–122

Tan L, Qu M, Zhu Y, Peng C, Wang J, Gao D, Chen C (2020) ZINC TRANSPORTER5 and ZINC TRANSPORTER9 function synergistically in zinc/cadmium uptake. Plant Physiol 183(3): 1235–1249

Tantipaiboonwong P, Pintha K, Chaiwangyen W et al (2017) Anti-hyperglycaemic and anti-hyperlipidaemic effects of black and red rice in streptozotocin-induced diabetic rats. Sci Asia 43(5):281–288

Thangadurai D, Kordrostami M, Islam S, Sangeetha J, Al-Tawaha ARMS, Jabeen S (2020) Genetic enhancement of nutritional traits in rice grains through marker-assisted selection and quantitative trait loci. In: Roychoudhury A (ed) Rice research for quality improvement: genomics and genetic engineering. Springer, Singapore, pp 493–507

Tharanathan RN (2002) Food-derived carbohydrates – structural complexity and functional diversity. Crit Rev Biotechnol 22:65–84. https://doi.org/10.1080/07388550290789469

Thomson MJ, Singh N, Dwiyanti MS, Wang DR, Wright MH, Perez FA, DeClerck G, Chin JH, Malitic-Layaoen GA, Juanillas VM, Dilla-Ermita CJ (2017) Large-scale deployment of a rice 6 K SNP array for genetics and breeding applications. Rice 10(1):1–13

Toreti A, Bassu S, Ceglar A, Zampieri M (2019) Climate change and crop yields. In: Ferranti P, Berry EM, Anderson JR (eds) Encyclopedia of food security and sustainability. Elsevier, Oxford

Trijatmiko KR, Duenãs C, Tsakirpaloglou N et al (2016) Biofortified indica rice attains iron and zinc nutrition dietary targets in the field. Sci Rep 6(1):1–13

Tripathy SK, Dash M (2021) Evaluation of rice germplasm–a quest for identification of high yielding iron rich rice. J Asian Agri 1:17–27

Tripathy SK, Patra S (2021) Elemental analysis of popular zinc rich rice varieties using ESEM and ICP-OES. Biosci Biotechnol Res Asia 18(2):305–311

Tripathy SK, Dash M, Behera SK et al (2017) Nutrient rich quality rice-a journey to healthy life. Adv Plants Agric Res 7:364–367

Umadevi M, Pushpa R, Sampathkumar KP, Bhowmik D (2012) Rice-traditional medicinal plant in India. J Pharmacogn Phytochem 1:6–12

Umadevi K, Niharika G, Vijayalakshmi V, Sanjeev D, Sheshu madav M (2019) Grain protein fractions of indian rice varieties. The Journal of Research, PJTSAU 47(2). https://epubs.icar.org.in/index.php/TJRP/article/view/92542

Valarmathi R, Raveendran M, Robin S, Senthil N (2015) Unraveling the nutritional and therapeutic properties of 'Kavuni' a traditional rice variety of Tamil Nadu. J Plant Biochem Biotechnol 3: 305–315. https://doi.org/10.1007/S13562-014-0274-6

Vasconcelos M, Datta K, Oliva N, Khalekuzzaman M, Torrizo L, Krishnan S, Oliveira M, Goto F, Datta SK (2003) Enhanced iron and zinc accumulation in transgenic rice with the ferritin gene. Plant Sci 164(3):371–378

Velprabakaran S, Rajeswari S, Jeyaprakash P, Raveendran M (2020) Genetic exploration on crude protein estimation in traditional rice landraces of southern India. Int J Curr Microbiol Appl Sci 9: 1072–1078. https://doi.org/10.20546/ijcmas.2020.904.127

Wada T, Uchimura Y, Ogata T, Tsubone M, Matsue Y (2006) Mapping of QTLs for physicochemical properties in japonica rice. Breed Sci 56:253

Wahyuni AS, Munawaroh R, Da' M (2016) Antidiabetic mechanism of ethanol extract of black rice bran on diabetic rats. https://doi.org/10.5455/njppp.2015.5.1111201590

Wang C, Han B (2022) 20 years of rice genomics research: from sequencing, functional genomics to quantitative genomics. Mol Plant 15(4):593–619

Wang L, Zhong M, Li X et al (2008) The QTL controlling amino acid content in grains of rice (*Oryza sativa*) are co-localized with the regions involved in the amino acid metabolism pathway. Mol Breed 21:127–137

Wang KM, Wu JG, Li G et al (2011) Distribution of phytic acid and mineral elements in three indica rice *(Oryza sativa* L.) cultivars. J Cereal Sci 54:116–121

Wang X, Song Y, Li J (2013) High expression of tocochromanol biosynthesis genes increases the vitamin E level in a new line of giant embryo rice. J Agric Food Chem 61:5860–5869. https://doi.org/10.1021/jf401325e

Wang DR, Agosto-Pérez FJ, Chebotarov D, Shi Y, Marchini J, Fitzgerald M, McNally KL, Alexandrov N, McCouch SR (2018) An imputation platform to enhance integration of rice genetic resources. Nat Commun 9(1):1–10

Wang S, Yang Y, Guo M, Zhong C, Yan C, Sun S (2020) Targeted mutagenesis of amino acid transporter genes for rice quality improvement using the CRISPR/Cas9 system. Crop J 8(3): 457–464

Wang Y, Meng Y, Ma Y et al (2021) Combination of high Zn density and low phytic acid for improving Zn bioavailability in rice (*Oryza stavia* L.) grain. Rice 14:23. https://doi.org/10.1186/s12284-021-00465-0

Welsch R, Li L (2022) Golden Rice – lessons learned for inspiring future metabolic engineering strategies and synthetic biology solutions. Methods Enzymol 671:1–29. https://doi.org/10.1016/BS.MIE.2022.03.014

Welsch R, Arango J, Bär C et al (2010) Provitamin a accumulation in cassava (Manihot esculenta) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. Plant Cell 22: 3348–3356. https://doi.org/10.1105/TPC.110.077560

Wen T-N, Luthe DS (1985) Biochemical characterization of rice glutelin. Plant Physiol 78:172–177. https://doi.org/10.1104/PP.78.1.172

Wenefrida I, Utomo HS, Linscombe SD (2017) Development and registration of 'Frontiere', a high-protein rice cultivar. J Plant Regist 11:240–244

Wickert E, Schiocchet MA, Noldin JA et al (2014) Exploring variability: new Brazilian varieties SCS119 Rubi and SCS120 Onix for the speciality rice market. Open J Genet 2014(4). https://doi.org/10.4236/ojgen.2014.42016

Wirth J, Poletti S, Aeschlimann B, Yakandawala N, Drosse B, Osorio S, Tohge T, Fernie AR, Gunther D, Gruissem W et al (2009) Rice endosperm iron biofortification by targeted and synergistic action ofnicotianamine synthase and ferritin. Plant Biotechnol J 7:631–644

Wong HW, Liu Q, Sun SSM (2015) Biofortification of rice with lysine using endogenous histones. Plant Mol Biol 87:235–248

Wu TY, Gruissem W, Bhullar NK (2018) Facilitated citrate-dependent iron translocation increases rice endosperm iron and zinc concentrations. Plant Sci 270:13–22

Wu TY, Gruissem W, Bhullar NK (2019) Targeting intracellular transport combined with efficient uptake and storage significantly increases grain iron and zinc levels in rice. Plant Biotechnol J 17 (1):9–20

Xu F, Sun C, Huang Y, Chen Y, Tong C, Bao J (2015) QTL mapping for rice grain quality: a strategy to detect more QTLs within sub-populations. Mol Breed 35:105

Xu M, Zhao S, Zhang Y et al (2017) Production of marker-free transgenic rice (*Oryza sativa* L.) with improved nutritive quality expressing AmA1. Iran J Biotechnol 15:102–110. https://doi.org/10.15171/ijb.1527

Xu Y, Wang X, Ding X, Zheng X, Yang Z, Xu C, Hu Z (2018) Genomic selection of agronomic traits in hybrid rice using an NCII population. Rice 11(1):1–10

Xu L, Yuan S, Wang X, Yu X, Peng S (2021a) High yields of hybrid rice do not require more nitrogen fertilizer than inbred rice: a meta-analysis. Food Energy Secur 10(2):341–350

Xu Y, Ma K, Zhao Y, Wang X, Zhou K, Yu G, Li C, Li P, Yang Z, Xu C, Xu S (2021b) Genomic selection: a breakthrough technology in rice breeding. Crop J 9(3):669–677

Yadava DK, Choudhury PR, Hossain F, Kumar D, Mohapatra T (2020) Biofortified varieties: sustainable way to alleviate malnutrition, 3rd edn. Indian Council of Agricultural Research, New Delhi, p 86

Yamada T, Matsuda F, Kasai K et al (2008) Mutation of a rice gene encoding a phenylalanine biosynthetic enzyme results in accumulation of phenylalanine and tryptophan. Plant Cell 20: 1316–1329

Yang R, Sun C, Bai J et al (2012) A putative gene sbe3-rs for resistant starch mutated from sbe3 for starch branching enzyme in rice (*Oryza sativa L.*). PLoS One 7:e43026. https://doi.org/10.1371/journal.pone.0043026

Yang A, Li Y, Xu Y, Zhang WH (2013) A receptor-like protein RMC is involved in regulation of iron acquisition in rice. J Exp Bot 64(16):5009–5020

Yang Y, Guo M, Li R, Shen L, Wang W, Liu M et al (2015) Identification of quantitative trait loci responsible for rice grain protein content using chromosome segment substitution lines and fine mapping of qPC-1 in rice (*Oryza sativa* L.). Mol Breed 35:130

Yang A, Li Q, Chen L, Zhang WH (2020) A rice small GTPase, Rab6a, is involved in the regulation of grain yield and iron nutrition in response to CO2 enrichment. J Exp Bot 71(18):5680–5688

Yano K, Yamamoto E, Aya K, Takeuchi H, Lo PC, Hu L, Yamasaki M, Yoshida S, Kitano H, Hirano K, Matsuoka M (2016) Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. Nat Genet 48(8):927–934

Yoo SC (2017) Quantitative trait loci controlling the amino acid content in rice (*Oryza sativa* L.). J Plant Biotechnol 44:349

Yu YH, Li G, Fan YY, Zhang KQ, Min J, Zhu ZW et al (2009) Genetic relationship between grain yield and the contents of protein and fat in a recombinant inbred population of rice. J Cereal Sci 50:121

Yu H, Xie W, Li J, Zhou F, Zhang Q (2014) A whole-genome SNP array (RICE 6 K) for genomic breeding in rice. Plant Biotechnol J 12(1):28–37

Yu X, Yang T, Qi Q et al (2021) Comparison of the contents of phenolic compounds including flavonoids and antioxidant activity of rice (*Oryza sativa*) and Chinese wild rice (*Zizania latifolia*). Food Chem 344:128600. https://doi.org/10.1016/j.foodchem.2020.128600

Yu H, Yang Q, Fu F, Li W (2022a) Three strategies of transgenic manipulation for crop improvement. Front Plant Sci 13:948518

Yu S, Ali J, Zhou S, Ren G, Xie H, Xu J, Yu X, Zhou F, Peng S, Ma L, Yuan D, Li Z, Chen D, Zheng R, Zhao Z, Chu C, You A, Wei Y, Zhu S, Gu Q, He G, Li S, Liu G, Liu C, Zhang C, Xiao J, Luo L, Li Z, Zhang Q (2022b) From Green Super Rice to green agriculture: reaping the promise of functional genomics research. Mol Plant 15:9–26

Yuyu C, Aike Z, Pao X, Xiaoxia W, Yongrun C, Beifang W, Yue Z, Liaqat S, Shihua C, Liyong C, Yingxin Z (2020) Effects of GS3 and GL3. 1 for grain size editing by CRISPR/Cas9 in rice. Rice Sci 27(5):405–413

Zaghum MJ, Ali K, Teng S (2022) Integrated genetic and omics approaches for the regulation of nutritional activities in rice (*Oryza sativa* L.). Agriculture 12(11):1757

Zhang Q (2021) Purple tomatoes, black rice and food security. Nat Rev Genet 22(7):414–414

Zhang X, Zhang G, Guo L, Wang H, Zeng D, Dong G et al (2011) Identification of quantitative trait loci for Cd and Zn concentrations of brown rice grown in Cd-polluted soils. Euphytica 180:173

Zhang Y, Xu YH, Yi HY, Gong JM (2012) Vacuolar membrane transporters OsVIT1 and OsVIT2 modulate iron translocation between flag leaves and seeds in rice. Plant J 72(3):400–410

Zhang GY, Liu RR, Xu G et al (2013) Increased α-tocotrienol content in seeds of transgenic rice overexpressing Arabidopsis γ-tocopherol methyltransferase. Transgenic Res 22:89–99. https://doi.org/10.1007/s11248-012-9630-2

Zhang P, Liu X, Tong H, Lu Y, Li J (2014) Association mapping for important agronomic traits in core collection of rice (*Oryza sativa* L.) with SSR markers. PLoS One 9(10):e111508

Zhang H, Shao Y, Bao J, Beta T (2015) Phenolic compounds and antioxidant properties of breeding lines between the white and black rice. Food Chem 172:630–639

Zhang C, Shinwari KI, Luo L, Zheng L (2018) OsYSL13 is involved in iron distribution in rice. Int J Mol Sci 19(11):3537

Zhang F, Wang C, Li M, Cui Y, Shi Y, Wu Z, Hu Z, Wang W, Xu J, Li Z (2021) The landscape of gene–CDS–haplotype diversity in rice: properties, population organization, footprints of domestication and breeding, and implications for genetic improvement. Mol Plant 14(5):787–804

Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat Commun 2(1):1–10

Zhao H, Yao W, Ouyang Y, Yang W, Wang G, Lian X, Xing Y, Chen L, Xie W (2015) RiceVarMap: a comprehensive database of rice genomic variations. Nucleic Acids Res 43(D1):D1018–D1022

Zheng Z, Sumi K, Tanaka K, Murai N (1995) The bean seed storage protein [beta]-phaseolin is synthesized, processed, and accumulated in the vacuolar type-II protein bodies of transgenic rice endosperm. Plant Physiol 109(3):777–786

Zheng X, Wu J-G, Lou X-Y, Xu H-M, Shi C-H (2008) QTL analysis of maternal and endosperm genomes for histidine and arginine in rice (*Oryza sativa* L.) across environments. Acta Agron Sin 34:369

Zheng L, Cheng Z, Ai C, Jiang X, Bei X, Zheng Y, Glahn RP, Welch RM, Miller DD, Lei XG et al (2010) Nicotianamine, a novel enhancer of rice iron bioavailability to humans. PLoS One 5: e10190

Zheng L, Zhang W, Chen X, Ma J, Chen W, Zhao Z et al (2011) Dynamic QTL analysis of rice protein content and protein index using recombinant inbred lines. J Plant Biol 54:321

Zheng L, Zhang W, Liu S, Chen L, Liu X, Chen X et al (2012) Genetic relationship between grain chalkiness protein content and paste viscosity properties in a backcross inbred population of rice. J Cereal Sci 56:153

Zhong M, Wang LQ, Yuan DJ, Luo LJ, Xu CG, He YQ (2011) Identification of QTL affecting protein and amino acid contents in rice. Rice Sci 18:187

Zhou J, Xin X, He Y, Chen H, Li Q, Tang X et al (2019) Multiplex QTL editing of grain-related genes improves yield in elite rice varieties. Plant Cell 38(4):475–485. https://doi.org/10.1007/s00299-018-2340-3

Zhu L, Gu M, Meng X, Cheung SC, Yu H, Huang J, Sun Y, Shi Y, Liu Q (2012) High-amylose rice improves indices of animal health in normal and diabetic rats. Plant Biotechnol J 10(3):353–362

Zhu Q, Yu S, Zeng D et al (2017) Development of "Purple Endosperm Rice" by engineering anthocyanin biosynthesis in the endosperm with a high-efficiency transgene stacking system. Mol Plant 10:918–929

Zhu C, Kobayashi K, Loladze I, Zhu J, Jiang Q, Xu X, Liu G, Seneweera S, Ebi KL, Drewnowski A, Fukagawa NK (2018) Carbon dioxide (CO_2) levels this century will alter the protein, micronutrients, and vitamin content of rice grains with potential health consequences for the poorest rice-dependent countries. Sci Adv 4(5):eaaq1012

Wheat Nutraceutomics: Breeding, Genomics, Biotechnology, and Nanotechnology

Velu Govindan, Om Prakash Gupta, Sunil Kumar, Chandra Nath Mishra, and Gyanendra Singh

Contents

V. Govindan (✉)
International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico
e-mail: velu@cgiar.org

O. P. Gupta · S. Kumar · C. N. Mishra · G. Singh (✉)
ICAR-Indian Institute of Wheat and Barley Research, Karnal, India
e-mail: director.iiwbr@icar.gov.in

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_2

Abstract

Wheat is humanity's second most important cereal crop, consumed widely in developed and developing countries and constituting a major source of protein and energy, especially in the developing world. Wheat grain, including the bran and endosperm, furnishes diverse macro-and micronutrients required for the normal physiological and biochemical functioning of the human body. Dietary deficiencies of micronutrients such as iron (Fe) and zinc (Zn) lead to severe health consequences in children below 5 years of age and in pregnant women and lactating mothers. Increasing the nutritional value of wheat grain can largely address micronutrient malnutrition for the world's growing population. The recent availability of wheat genome sequence library in the public domain, together with the expanding horizon of next-generation sequencing, and genome editing technologies, holds great promise for trait-based molecular breeding to develop nutrient-rich wheat cultivars. Modern biofortification techniques, including conventional breeding, transgenics, and agronomic biofortification, have already increased wheat grain nutrient content and the nutrient-rich biofortified wheat cultivars grown over 2 million ha area in South Asia & Latim America. This chapter discusses the importance of wheat in the human diet, wheat grain's nutritional composition, and advances in molecular and transgenic and genome editing approaches to develop health-related traits in wheat grain.

Keywords

Micronutrients · Malnutrition · Genome editing · Genomic selection · QTLs · Genetic engineering

1 Introduction

1.1 The Importance of Wheat

Wheat, an annual herb belonging to the family Gramineae or Poaceae, is a food crop grown and consumed in nearly 100 countries and imported and consumed in many others where western-style diets are being adopted (Cummins and Roberts-Thomson 2009; Shewry 2009). Common hexaploid wheat (*Triticum aestivum* L., 2n = 6x = 42, AABBDD) is one of the most important staple crops in the world, serving as a key food source for 30% of the human population and contributing approximately 20% of its energy needs (calories) and 25% of its dietary protein (Borisjuk et al. 2019). Its grain is used in bread, supporting the baking industry, and popular foods such as chapatis, noodles, and cookies, to name just a few. Worldwide wheat cultivation ranges from 67° N in Scandinavia and Russia to 45° S in Argentina, including elevated regions in the tropics and sub-tropics (Feldman 1995; Shewry 2009). About 95% of wheat grown worldwide is hexaploid bread wheat (*Triticum aestivum* L.), with most of the rest being tetraploid durum wheat (*T. turgidum* var. *durum*), which

is adapted to dry, Mediterranean climates and used in pasta. Minor amounts of primitive wheats are also grown mainly for specialty health foods, including einkorn (diploid *Triticum monococcum*), emmer (tetraploid *T. turgidum* var. *dicoccon*), and spelt (hexaploid *T. aestivum* var. *spelta*) in certain areas of Spain, Turkey, the Balkans, and the Indian subcontinent. The latter differs from bread wheat essentially in that the hull is not removed by threshing, resulting in a higher fiber content when consumed as whole grain (Brouns et al. 2013).

Cultivation of wheat started about 10,000 years ago as part of the Neolithic Revolution, when humans switched from hunting and gathering to settled agriculture. The earliest cultivated forms were diploid (einkorn with genome AA) and tetraploid (emmer with genome AABB) wheats from southeastern Turkey (Brouns et al. 2013; Dubcovsky and Dvorak 2007). Hexaploid bread wheat (AABBDD) is believed to have emerged some 9000 years ago through spontaneous hybridization between a cultivated tetraploid (*Triticum turgidum;* AABB) and goat grass (*Aegilops tauschii;* DD) (Brouns et al. 2013; Feldman 2001). The earliest cultivated forms were landraces presumably selected from wild populations by ancient farmers, considering their superior yield and other agronomical important characteristics, a domestication that separated modern wheat genetically and phenotypically from its wild relatives and early forms.

Wheat grain production amounted to over 780 million tons (t) harvested from more than 225 million hectares (ha) in 2019–2020 (http://www.fao.org/faostat); but wheat is still the third major food crop, lagging behind maize and rice both in yield and the application of genomic tools for crop improvement (Borisjuk et al. 2019; Uauy 2017). Average wheat yield worldwide increased nearly three-fold during the Green Revolution of the mid-to-late twentieth century, largely due to expanded irrigation, intensive fertilizer application, and advanced breeding methods (Evenson and Golin 2003a), but the current global average yield around 3 t/ha is far below the crop's genetic potential (Langridge 2013), aside from yield gaps relating to crop management. As estimated by Langridge (2013) and Henry et al. (2016), to meet the wheat consumption demands (expected to rise 1.6% annually) of an estimated 9.5 billion world population by 2050, wheat yields should grow by over 60% to approximately 5 t/ha, coupled with maintaining or improving its nutritional characteristics and using currently available land. Facing this challenging scenario, which includes rising temperatures and alarming water scarcities, the emphasis must be on improved productivity and adapting to environmental challenges (Borisjuk et al. 2019).

1.2 Wheat's Importance in Times of Chronic Disease and Malnutrition

Malnutrition can be classified as under-nutrition (hunger, micronutrient malnutrition/hidden hunger) and over-nutrition (overweight/obesity). For malnourished children under 5 years of age, 149 million are stunted, 49.5 million are wasted, and 40 million are overweight (Development Initiatives 2020; Poole et al. 2021). About 45% of

mortality among children aged five and below is associated with malnutrition, chiefly in low- and middle-income countries. In the early 2000s, more than 3 billion people were micronutrient malnourished (Šramková et al. 2009). More recently, some 528 million (29%) women of reproductive age are anemic from a lack of dietary iron, making iron deficiency the most widespread micronutrient deficiency in the world (Choge 2020). Dietary deficiencies of zinc, a critical micronutrient, are widespread, accounting for malnutrition-related developmental impediments across all age groups. Lethal effects amount to 800,000 child deaths annually, with vulnerability concentrated in sub-Saharan Africa and South Asia. Staple cereals such as wheat are significant sources of both minerals, contributing 44% of the daily intake of iron (15% in bread) and 25% of the daily intake of zinc (11% in bread) in the UK (Henderson et al. 2007; Shewry 2009). According to an estimate, almost 690 million people suffered from hunger in 2019, worsened by the worldwide COVID-19 health pandemic. The number of hungry people is expected to exceed 840 million by 2030 – almost 10% of the global population (Poole et al. 2021). While humankind is fighting malnutrition on one front, over-nutrition is increasing globally: nearly half of the world's adult population is overweight or obese, and three-quarters of those persons live in low- and middle-income countries (Poole et al. 2021). Worldwide incidences of diabetes (44%), ischemic heart disease (23%), and certain cancers (7–41%) are linked to being overweight and obese.

Child stunting often results from micronutrient malnutrition tied to imbalanced diets in children and mothers, especially among the poor. Food-based approaches to prevent malnutrition and which focus on micronutrients can help address "hidden hunger." Cereal-based allergies and intolerance have also posed serious concerns. Major chronic diseases include obesity, heart ailments, cancer, and celiac disease, among others, taking a heavy toll on health and the world economy. In the current context of chronic diseases, wheat, its bioactives, and products assume relevance and need to be deliberated in a multifarious context.

1.3 The Limitations of Conventional Breeding and Rational for Next-Generation Breeding: Nutritional Perspectives

Wheat underwent hybridization and genome duplication to generate its hexaploid genome ($2n = 6x = 42$, AABBDD). Bread wheat possesses a sesquipedalian genome – 17 gigabases – which is over 5 times larger than the human genome and 40 times the size of the rice genome. Recent estimates document 107,891 high-confidence genes in bread wheat, with over 85% repetitive DNA sequences, representing a three-fold redundancy associated with being hexaploid. Every year, conventional commercial breeding produces a large number of new crop varieties to improve productivity and nutrition, strengthen food security, and increase consumer acceptability (Govindan et al. 2022). The conventional breeding process has evolved to provide an effective framework for improving crop performance while also assisting in the development of safe and nutritious foods. Conventional plant breeding entails identifying desirable parents in order to create favorable combinations in the next generation (Kaiser et al. 2020). This selection of a few individuals

from a large population is an important component of the plant breeding process (Kaiser et al. 2020). Conventional breeding has made significant contributions to large-scale cultivation, yield potential, and the frequency of desirable traits in wheat; however, it has certain boundaries that should be considered under current conditions, and a new wheat breeding strategy is required (Borisjuk et al. 2019).

Wheat's hexaploid genome and associated functional gene redundancy make genetic advances to selecting a desired phenotype difficult, if not impossible, due to gene linkage or gene drag. Further advancement in wheat breeding is dependent on understanding of functional genomics. Grain yield and quality can be improved by identifying the most important key genes, as well as their structures, roles, and functions in wheat plant development. This functional genomics knowledge can then be used to change the structures and functions of selected key genes via genetic manipulation (Borisjuk et al. 2019), a broad term used here to describe molecular methods whose products fall outside the traditional definition of "genetically modified" (GM) such as RNA interference (RNAi) and clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein (CRISPR/Cas9) Borisjuk et al. 2019). Such approaches will be critical in determining the functions of wheat genes. By using cutting-edge technology like genome sequencing and targeted mutations using genome editing methods like CRSPR-Cas and RNAi, it is possible to increase the resilience of wheat while reducing environmental pollution (Gupta and Karkute 2021). However, the use of CRISPR/Cas systems depends on knowledge of the targeted gene's sequence. Through homology directed repair (HDR), precise substitution of an existing allele has substantially benefited crop improvement with elite alleles in commercial types. Base editing, prime editing (PE), genome sequencing, genome-wide association study (GWAS), speed breeding, high-throughput genotype and phenotype profiling, and synthetic biology are a few examples of contemporary methods (Gupta et al. 2022). The ability to accurately replace one base with another by base editing, an alternative and powerful method for HDR-mediated gene replacement, has made it possible to accurately replace an allele with a single nucleotide polymorphism (SNP) (Komor et al. 2016). Many base substitutions (12 types) and minor insertions-deletions (indels) are made possible by prime editing (PE), which increases the reach and potential of precision genome editing. Recent studies have shown that RNAi, a frequent mechanism for controlling gene expression in eukaryotic cells, is a reliable tool for functional genomics and the engineering of novel phenotypes. The method relies on the expression of small interfering RNA (siRNA) molecules, such as antisense or hairpin RNAi constructs, to affect post transcriptional gene silencing in a sequence-specific manner (Borisjuk et al. 2019). RNAi applications in wheat.

2 Detailed Nutritional Composition of the Wheat Grain

Grain is the harvested and economically important part of the wheat plant, and its biochemical composition determines its nutritional and health properties (Shewry 2009). Table 1 details the generalized composition of various chemical constituents. On a dry weight basis, a wheat grain can be divided into three distinct parts: mealy

Table 1 Contents and variations in chemical composition of wheat grains (dry weight)

Parameter	Range (%)	References
Carbohydrates (%)	85	Stone and Morell (2009)
Starch (20–30% amylose and 70–80% amylopectin)	55–75	Borisjuk et al. (2019)
Dietary fiber/non-starch/cell wall polysaccharides (%)	11.5–15.5	Andersson et al. (2013)
Fats (%)	2.0–2.5	Gonzalez-Thuillier et al. (2015)

endosperm (80–85%), outer bran (13–17%), and germ (2–3%). (Belderok et al. 2000). The endosperm occupies the majority of the wheat grain, primarily consisting of starch and proteins with small amounts (2%) of fiber. The endosperm provides energy to the seed, making available carbohydrates and proteins during germination. The germ, the smallest segment of grain, contains lipids, sterols, antioxidants, vitamins (B, E), minerals, and enzymes, as well as nourishing the seed. The bran is a seed's outer shell that contains fiber, vitamins, and trace minerals. The outer pericarp (3–5%) contains insoluble dietary fibers and bound phenolic acids (which act as antioxidants), whereas the aleurone layer (6–9%) contains dietary fibers, proteins, enzymes, phenolic compounds, vitamins (B-complex), minerals, and phytates. The testa (1% of the total) contains trace amounts of alkylresorcinols, sterols, and steryl ferulates. Wheat bran may typically contain dietary fiber (42.8%), other carbohydrates (21.7%), protein (15.6%), ash (5.8%), and lipids (4.3%).

There is a need for suitable and efficient methods to test diverse wheat genotypes for various quality parameters, as documented in the updated listing by Gupta et al. (2022). Model-based methodologies are being developed to identify superior wheat lines in early generations of breeding cycles (Mohan et al. 2022).

2.1 Carbohydrates

Carbohydrates account for the majority of wheat grain (up to 85%) (Stone and Morell 2009). Starch accounts for the majority of stored carbohydrates, accounting for 55–75% of grain dry weight (Borisjuk et al. 2019). Fibers and low-molecular-mass mono-, di-, and oligo-fructans are examples of others (Table 1). Wheat grain starch is found in either large lenticular granules of 25–40 m, which develop during the first 15 days after pollination, or small spherical granules of 5–10 m, which develop 10–30 days after pollination and account for approximately 88% of total grain starch granules (Belderok et al. 2000). Starch is a polymeric form of glucose that is chemically classified as amylose and amylopectin. Amylose has a molecular weight of around 250,000 and may contain nearly 1500 glucose units with wide variations. Amylose is thought to have a linear polymer structure, with -(1,4)-glycosidic linking glucose moieties together and a degree of polymerization (DP) of 1000–5000 glucose units. The structure of this polymer was previously assumed

to be primarily linear, but this appears to be true for only a portion of the amylose; the remainder is slightly branched. The branching in amylopectin is to a much greater extent than in amylose and the average unit chain of amylopectin has only 20–25 glucose molecules, with an average molecular weight of about 10^8. Amylopectin is a much larger polymer with a DP ranging from 105 to 106 glucose units tethered via α-(1,4)-linked glucose polymers, which are further connected by α-(1,6)-linkages (5–6%).

2.2 Dietary Fiber

Dietary fiber is defined as lignin plus plant polysaccharide components that are indigestible by human digestive enzymes. While soluble fiber (pectic substances, hydrocolloids, -glucans) is water soluble, insoluble fiber (cellulose, hemicellulose, lignin, arabinixylans) is not. Regular consumption of dietary fiber, which is primarily found in whole grains, protects against heart disease, hypertension, hyperlipidemia, type 2 diabetes, obesity, constipation prevention, diverticular disease, esophageal disease, and a variety of cancers (Poole et al. 2021; Weickert and Pfeiffer 2008). Soluble fiber such as β-glucan [(1→3,1→4)-β-D-glucan] effects glycemic index and appears to help prevent chronic diseases like diabetes and obesity and, possibly, negative effects associated with FODMAPS (fermentable oligo-, di-, monosaccharides and polyols) (Poole et al. 2021). Arabinoxylans (AX) and (1→3), (1→4)-β-glucans are housed primarily in wheat endosperm cell walls. Arabinoxylans yield short-chain fatty acids, particularly butyrate, in the colon. It is conjectured that high butyrate concentrations in the colon improves bowel health and lower cancer risk.

2.3 Proteins

Proteins are required for various body functions, ranging from enzymatic, structural to locomotory and many more. Protein content in wheat grain ranges from 10 to 15% (dry wt.) (Borisjuk et al. 2019). Wheat proteins are classified according to their extractability and solubility as per Osborne and have been highlighted in Table 2 In general, cereal proteins are low in the essential amino acids lysine (1.5–4.5% vs. 5.5% of WHO recommendation), tryptophan (Trp, 0.8–2.0% vs. 1.0%), and threonine (Thr, 2.7–3.9% vs. 4.0%).

2.4 Lipids

Lipids are present in small amounts (2–2.5%) in wheat but have a significant impact on food quality and texture because they can bind to proteins and starch to form inclusion complexes. The germ contains nearly 11% of total lipids, but significant amounts are also associated with the endosperm's bran, starch, and proteins

Table 2 Types of wheat proteins with their solubility and other features

Protein name	Solubility	% of Total protein	Features
Albumins	Water soluble	Smallest in size (10%)	Most of physiologically active enzymes belong to these two; both proteins are present in the seed coat, the aleurone cells and the germ, with a lower concentration in the mealy endosperm; both make up 20–25% of total wheat proteins
Globulins	Insoluble in pure water; soluble in dilute NaCl solutions, but insoluble at high NaCl concentrations	Size more than albumins (10%)	
Gliadins	Soluble in 70% ethyl alcohol	Size more than above two (45%)	High-molecular-weight storage proteins for future use by the seedling located in the mealy endosperm; both constitute 75–80% of total wheat proteins; both are unique, being biologically active: though they have no enzyme activity, they function in dough formation through gas retention, producing spongy baked products
Glutenins	Soluble in dilute acid or sodium hydroxide solutions	Low- and high-molecular-weight types (35%)	

Source: Belderok et al. (2000), Borisjuk et al. (2019), Šramková et al. (2009)

(Poole et al. 2021). All fractions contain significant amounts of free fatty acids and triacylglycerols (Gonzalez-Thuillier et al. 2015). Phosphatidyl choline, phosphatidyl ethanolamine, and phosphatidyl serin are the most common bound lipids, followed by lysophosphatidyl derivates with one free hydroxyl group on the glycerol moiety. The main sterols were identified as -sitosterol, campesterol, and saturated sterols C28 and C29. Numerous studies have shown a high level of linoleate (C18:2) in both the total lipid and the triglycerides.

2.5 Vitamins

Over 3 billion people are currently micronutrient malnourished, which means their diets are deficient in micronutrients such as vitamins. Vitamins are a diverse group of food-based, essential, small organic substances that are synthesized by plants and microorganisms rather than the human body. They do not provide energy but are essential micronutrients for humans, acting as coenzymes or their precursors (niacin, thiamin, biotin, pantothenic acid, vitamin B6, vitamin B12, and folate) or in specialized functions such as vitamin A in vision and ascorbate in specific hydroxylation reactions. Vitamins play roles in human genetic regulation and genomic stability (folic acid, vitamin B12, vitamin B6, niacin, vitamin C, vitamin E, and vitamin D) as well as antioxidative defense systems (vitamins C and E and some carotenoids) (Poole et al. 2021). The vitamins in wheat and their range are presented in Table 3.

Table 3 Vitamins, phenolics, minerals, and anti-nutritional factors composition in wheat grains

Major component	Name	Range	References
Vitamins	Vitamin B1 (Thiamine) (mg/kg)	5.53–13.55	Shewry et al. (2011)
	Vitamin B2 (Riboflavin) (mg/kg)	0.77–1.40	
	Vitamin B3 (Niacin) (mg/kg)	0.16–1.74	
	Vitamin B5 (Pantothenic acid) (mg/kg)	0.88–4.04 (Durum)	Tekin et al. (2018)
	Vitamin B6 (Pyridoxine) (mg/kg)	1.44–3.05	Batifoulier et al. (2006)
	Vitamin B9 (Folates) (mg/kg)	0.323–0.774	Piironen et al. (2008)
Minerals	Magnesium (mg/kg)	600–1400	Oury et al. (2006)
	Iron (mg/kg)	18–40	Ram and Govindan (2020)
	Zinc (mg/kg)	21–63	
	Copper (mg/kg)	1.8–6.2	
	Manganese (mg/kg)	24–37	
Phenolics	Phenolic acids (μg/g)	326–1171	
Anti-nutritional factors	Phytic acid (mg/g)	12–18	

2.6 Tocols

The germ fraction of t einkorn accessions and some bread wheat seed showed the highest concentrations of -tocopherol and total tocols. The bran fraction had the highest levels of -tocotrienol, but significant amounts were also found in the flour.

2.7 Minerals

Table 3 summarizes the minerals that are available in wheat, as well as their generalized availability ranges. The topic of two vital minerals, iron and zinc, has been addressed here. Plants primarily store iron in the form of ferritin structures, which accumulate primarily in non-green plastids, etioplasts, and amyloplasts (Borisjuk et al. 2019). Iron has been found primarily in the aleurone layer (bran) of wheat grains, where it has complexed with phytate (myo-inositolphosphate 1,2,3,4,5,6-hexa-kisphosphate). These complexes are insoluble, limiting iron bio-availability in humans and livestock. Scientists are experimenting through breeding to express phytase enzymes in developing grain and thus increase mineral availability. The discovery of a heat-stable form of this enzyme would allow phytate complex hydrolysis to occur during food processing. Another option is to increase the concentration of Fe in grains. In 1994, Fe concentrations in a wheat variety grown at the CIMMYT research station in El Batán, Mexico, ranged from 28.8 to 56.5 mg/kg (mean = 37.2 mg/kg). Peleg et al. (2008) reported new wild emmer wheat accessions with very high Fe (up to 88 mg/kg) and Zn (up to 139 mg/kg)

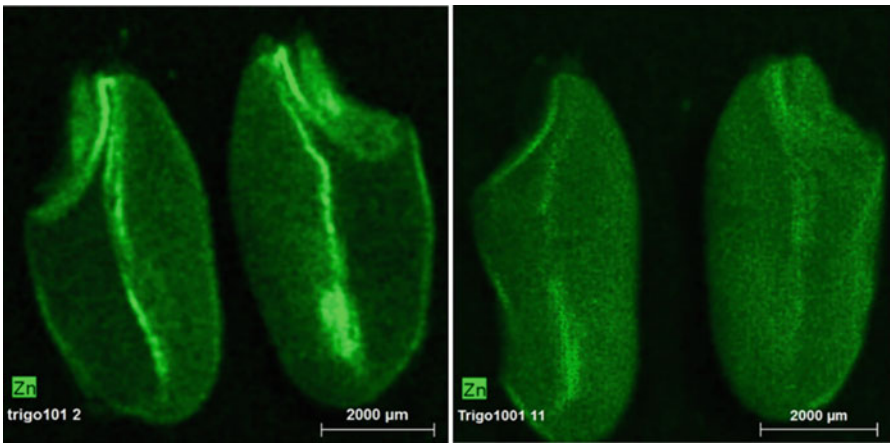


Fig. 1 Localization of zinc in wheat grain with µXRF (left high zinc wheat 'Zinc-shakti', right CIMMYT control variety 'Baj')

concentrations, as well as high protein content (up to 380 g/kg) and tolerance to drought and Zn-deficient soils.

Zn deficiency causes nearly 500,000 deaths in children under the age of five each year (Borisjuk et al. 2019). Micronutrient levels in modern elite wheat cultivars are typically suboptimal. Given the high concentrations of Zn and Fe in the outer husk, aleurone, and embryo, both micronutrients are lost during milling and polishing. Phytate, an anti-nutritional factor, reduces micronutrient availability in the human digestive tract. Zn deficiency affects nearly 33% of the world's population, resulting in health complications such as stunted physical development, weakened immunity, and decreased learning ability, among other things. According to a variety of reports and survey studies, the average Zn concentration in whole wheat grain is from 21 to 63 mg/kg (Ram and Govindan 2020). Most of wheat grain Zn is in the embryo and aleurone layer, with a small portion in the endosperm (Fig. 1).

3 Marker-Assisted Breeding for Health-Related Traits

Genetic tools can accelerate genetic gains for yield and stress resilience by assisting in precise and accurate selection, saving time, money, and labor in crop improvement. Breeding approaches combine modern cutting-edge genomic tools and breeding tools to accelerate breeding progress. Molecular marker-assisted selection (MAS), backcross breeding (MABB), and recurrent selection (MARS) aid in the early identification of favorable alleles for economically important traits (Bonnett et al. 2005). It should also be noted that the requirements for selecting markers, as well as overestimation of marker effects with minor contributions, can limit the effectiveness of using molecular markers in MAS. In wheat, MAS aids in the improvement of agriculturally important traits by allowing efficient screening of

difficult-to-estimate traits, the transfer of genomic regions from genetic stocks rich in desirable traits into better backgrounds, and the pyramiding of different polygenic characters. The discovery of numerous QTL-specific molecular markers for monogenic and polygenic traits in recent years has accelerated the deployment of molecular markers for regions associated with biotic and abiotic resistance, as well as other economically important traits. The cost of more precise genotyping, high-throughput genotyping and phenotyping, and the use of imaging and computational traits have all reduced the cost of biotechnological tools.

3.1 Germplasm Characterization

The high level of variability in nutritional traits found in wild relatives and land races of wheat allows for the development of high-yielding nutrient-rich wheat varieties (Cakmak 2008). The traits required for desirable processing and end-use quality, as well as nutritional traits, are being transferred from wild relatives of wheat, such as *Aegilops tauschii*, *Triticum turgidum* ssp. *diccocoides*, *Triticum turgidum* ssp. *dicoccum*, and *Triticum aestivum* ssp. *spelta* species, to high-yielding bread wheat lines that feature high yield and better adaptation (Guzmán et al. 2014). *T. dicoccon*-based synthetic hexaploids, landraces from Iran, Spain, and Afghanistan, and *T. dicoccoides* from Israel and adjacent regions are being used to improve micronutrients. CIMMYT Mexico has a large collection of genetic resources in its germplasm bank near Mexico city.

Screening for micronutrient variability has demonstrated that landraces and wild relatives of common wheat such as *T. spelta* and *T. dicoccon* have the ample amount of Zn and Fe. Contemporary breeding methods employ limited backcrossing procedure to introgress high-Zn/Fe genes from *T. spelta,* synthetic hexaploids, and landraces into better agronomy genotypes available with the breeders.

Ae. peregrina accessions have more than double grain Fe and Zn concentrations than elite wheat cultivars. Some of the derivatives of fertile wheat x *Ae. peregrina* with bolder seeds, better harvest indexes comparable to those of elite wheat lines, and higher micronutrient concentrations demonstrated that *Ae. peregrina* possesses a separate genetic system for biofortification, similar to that of *Ae. kotschyi*. The fertile BC_2F_2 plants with one or more additional chromosomes from *Ae. peregrina* showed a 100–200% enhancement of grain Fe and Zn levels over those of normal, elite wheat lines. Further analysis showed that the two chromosome groups 7 and 4 of *Ae. peregrina* had genomic regions for micronutrient content in wheat.

3.2 Marker-Assisted Gene Introgression

The use of MAS can make conventional wheat breeding programs more cost-effective and time-efficient (Gupta et al. 2010). It has primarily been used in wheat for foreground for carrier chromosome or segment selection and background selection for maximum genome recovery. It has recently been used successfully to

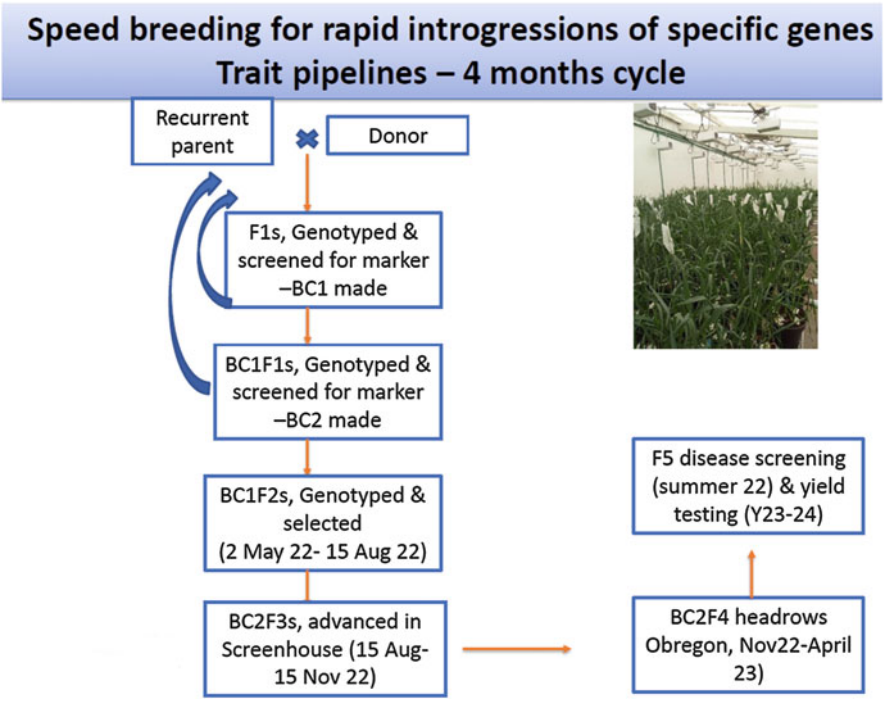


Fig. 2 Trait augmentation program at CIMMYT, Mexico

introduce and pyramid major genes/QTL for desirable wheat characteristics of a major locus (Gpc-B1) on chromosome 6BS that was introgressed from wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*) has improved Fe (18%), Zn (12%), and protein (38%). A marker called *Xuhw89*, which is closely related to the (0.1 cM) *Gpc-B1* locus (Distelfeld et al. 2006), has aided in the development of wheat varieties with high Fe, Zn, and protein concentrations. Many RFLP, SSR, and CAPS markers have been linked to this marker *Xuhw89* (Distelfeld et al. 2004). With the cloning and characterization of *Gpc B1* locus, a gene-specific marker and the locus has been introduced in wheat, improving GPC without a yield penalty (Kade et al. 2005). This has occurred mostly in many wheat growing regionssuch as India where it has transferred to elite cultivars. The wheat breeding program at CIMMYT recently embarked upon to introgress key agronomic, disease resistance and nutritional quality traits through its speed breeding pipeline (Fig. 2).

3.3 Gene Pyramiding

Functional markers are critical for gene stacking, genomic region transfer, and gene editing procedures. So far, 97 functional markers in bread wheat have been used to establish 30 loci from 93 alleles (Alotaibi et al. 2021). With the advancement of

genomic studies, the number of alleles has increased, and approximately 157 markers for 100 loci for various traits of economic importance have been identified. Few genes, including *VRN1* and *Gpc-B1*, have been successfully incorporated using the conventional positional cloning approach in wheat. Similarly, genes related to grains, such as *TaGS5*, *TaGS3*, and *TaCwi-A1* related to grain size; the *Psy1* (He et al. 2009) phytoene synthase gene and *Zds1* related to zeta-carotene desaturase have been reported using competitive genomics methods.

4 Identification, Cloning, and Characterization of Health-Related Genes/QTLs

Micronutrient effectiveness has a complex nature that is influenced by the environment. It is critical for MAS and map-based cloning to identify the genes/QTLs that influence micronutrient content. Map-based cloning is a method for cloning the gene of interest without prior knowledge of its product. The basic requirement for map-based cloning is a population that has been systematically developed for the desired trait and is suitable for fine mapping. Precision genotyping and phenotyping for the trait aid in the generation of an accurate genetic map indicating the location of the gene. To determine the physical location, two nearby markers are used to screen BAC libraries. The candidate genes are discovered through chromosome walking and target interval sequencing. Significant progress has been made since cloning of the first gene. The gene *Gpc-B1* has been positionally cloned in wheat; it is 7.4 kb (md) long and encodes for NAC transcription factors controlling senescence, protein content, and Zn and Fe content.

A major QTL (18.3% phenotypic variation) in rice for Fe content on chromosome 8 of rice and shows synteny with chromosome 7 of wheat. Similarly, in a *T. durum x T. dicoccoides* cross combination, Peleg et al. (2008) discovered a major locus Fe content on chromosome 4 and two co-located loci on chromosome 7 for both Fe and Zn. Discovered major QTL on chromosome 7A for both micronutrients in a *T. boeoticum* and *T. monococcum* population. Xu et al. (2012) discovered 9 additive and 4 epistatic QTL on the 4B and 5A chromosomes, indicating a shared inherent basis for the three nutritional traits. Populations of diploid wheats, durum wheats, and wild Emmer wheat (Peleg et al. 2009), as well as synthetic wheats and *T. spelta*, were used to map QTLs for grain Zn and Fe concentration. In a separate report, Srinivasa et al. (2014) identified 10 QTLs (five each for Zn and Fe accumulation) that were widely distributed across 7 chromosomes. In a DH (doubled haloid) population, two QTLs for Zn content were found on chromosomes 1B and 2B. In addition, four genes governing the inheritance of grain Zn content were discovered in two mapping populations derived from a *T. spelta* x bread wheat combination (Srinivasa et al. 2014). Identified several QTLs on chromosome 7B accounting for 32.7% of total phenotypic variation for Zn concentration and a single QTL location on 4A accounting for 21.14% of total phenotypic variation for grain Fe content in two RIL populations derived from *T. spelta* L. and synthetic hexaploid wheat crosses. The majority of studies have found a statistically significant positive

relationship between grain Zn and Fe concentrations in various environments where QTLs or similar genetic effects have regulated the Fe and Zn content in wheat. Co-localization of Fe and Zn concentrations has also been shown on chromosomes 2A, 2B, 4BS, and 5A (Xu et al. 2012) and 6B. This co-location of QTLs offers an opportunity to pursue a single marker-assisted program to enhance the concentrations of both Zn and Fe content. Table 4 provides a comprehensive list of major genes/QTLs identified in wheat for various quality traits.

5 Genomics-Aided Breeding for Heath-Related Traits

In wheat, genome-wide association studies (GWAS) have been widely used to investigate the genetics of quantitative traits. In comparison to traditional QTL mapping, GWAS provides better QTL resolution and wider allelic coverage, and it can be used on a larger population of genetic resources, landraces, varieties, or advance lines. Only a few studies in wheat have been conducted to better understand the genetic mechanism of quality characters. Nonetheless, improved genotyping facilities and access to the wheat genome reference sequence, RefSeq v.1.0, can speed up the identification and prediction of markers and their effects on trait. It would improve trait mapping, gene discovery. Detection of SNP markers that are distributed genome wide offers new avenues for genetic improvement of bread wheat for yield and traits of economic importance (Juliana et al. 2019).

GWAS and interval mapping studies in wheat have resulted in the identification of hundreds of markers for improving Fe and Zn concentration, but only a few have been used in marker-assisted breeding for nutritional traits. The only example is the *Gpc-B1* gene, which has been used to improve Fe, Zn, and Protein concentration in wheat using a MAS approach. Developed chromosome substitution lines in the variety 'Langdon' (LDN) and reported that a locus *QGpc.ndsu-6B* with a phenotypic variation of 66% is located on the 6B chromosome and contributes to high GPC. The wild allele at the locus accelerates leaf senescence and reduces grain size, resulting in yield reduction. *Gpc-B1* regulates senescence and nutrient remobilization, according to studies. Normally, wheat cultivars carry a non-functional *NAM-B1* allele, which is produced by a frame shift mutation in the wild allele, and this non-functional allele was preferred during wheat domestication. The presence of this allele on chromosome 6BS in wheat (Brevis and Dubcovsky 2010) allows for more time for grain development, which increases grain size and yield. In a study of 367 global bread wheat genotypes, only 5 Fennoscandian cultivars were found to have functional *Gpc-B1* or *NAM-B1* alleles, and these genotypes were only present for a short time in northern Europe. According to the wild-type/functional *Gpc-B1* allele has been conserved during the process of domestication. The MAS to transfer the wild-type *Gpc-B1* gene from Canadian to Australian varieties and found no yield penalty. Reported function loss caused by GPC1 and GPC2 mutations. GPC was found to be negatively associated with grain yield and influenced by genetic background in the majority of studies (Brevis and Dubcovsky 2010), and it is positively correlated with protein, iron, and zinc content with slight negative effect on yield. It is also suggested

Table 4 QTLs for Fe, Zn, and Selenium content in the grain of wheat and wild species

S. No	Cross/parents	Traits	QTLs	References
1.	*Triticum dicoccoides*	Fe & Zn	*GPC-B1* (6 7 BS)	
2.	*Triticum dicoccoides*	Fe & Zn	*TtNAM-B1*	
3.	*Hexaploid wheat (6x)* W7984 × Opata85			
4.	Hanxuan l0 × Lumai 14			
5.	RAC875-2 × Cascades			
6.	RIL (*Triticum boeoticum* × *Triticum monococcum*)	Fe	*QFe.pau-7A, QFe.pau-2A*	
7.	*Tetraploid wheat (4x)* Langdon × Accession #G18-16			Peleg et al. (2009)
8.	RIL (Xiaoyan × 54 Jing 411)	Fe & Zn	*QZn-5A, QFe-5A2, QGpc-5A1, QGpc-6A*	Xu et al. (2012)
9.	Tabassi × Taifun			
10.	*Tetraploid wheat (4x)* LDN × G18-16	Se		
11.	Berkut 9 × Krichauff			
12.	SHW L1 × Chuanmai32 & Chuanmai32 × Chuannong16	Fe, Zn & Se		Pu et al. (2014)
13.	RIL (PBW343 × Kenya Swara)	Zn	*QGzncpk.cimmyt-1BS, QGzncpk.cimmyt-2Bc, QGzncpk.cimmyt-3AL*	
14.	RIL (*T. spelta* (H+ 26 (PI348449) × *T. aestivum* cv. HUW 234)	Fe & Zn	*QZn.bhu-2B, QZn.bhu-6A, QFe.bhu-3B*	Srinivasa et al. (2014)
15.	DH (Berkut × Krichauff) Hexaploid (Adana99 × 70.711)	Fe & Zn	*QGfe.ada-2B, QGfe.ada-2B, QGZn.ada-2B, QGfe.ada-2B, QFe.bhu-2B*	
16.	*T. spelta* accession H + 26 (PI348449) × HUW 234			Srinivasa et al. (2014)
17.	*Hexaploid wheat (6x)* SHW-L1 × Chuanmai 32	Se		Pu et al. (2014)
18.	Seri M82 × SHW CWI76364			
19.	Tetraploid (Saricanak98 × MM5/4)	Fe & Zn	*QGfe.sar-5B& Qzneff.sar-6A, Qzneff.sar-6B& QGzn.sar-1B, QGzn.sar-6B, QGZn.sar-1B*	
20.	Saricanak98 × MM5/4 (4 × wheat)			Velu et al. (2017)
21.	DH (Berkut × Krichauff)	Zn	*QZn.bhu-1B, QZn.bhu-2*	
22.	Hexaploid (Adana99 × 70.711)	Zn	*QGzn.ada-6B, QGzn.ada-1D, QGzn.ada-7B*	Velu et al. (2017)

(continued)

Table 4 (continued)

S. No	Cross/parents	Traits	QTLs	References
23.	Adana99 × *T. Sphaerococcum* (70.711)			Velu et al. (2017)
24.	RIL (Synthetic hexaploid wheat × *Triticum spelta*)	Fe & Zn	*QGZn.cimmyt-7B_1P2, QGFe.cimmyt-4A_P2, QGZn.cimmyt-7B_1P2, QGZn.cimmyt-7B_1P1*	
25.	*Triticum dicoccon* PI94624/ *Aegilops squarrosa* [409] × BCN	Fe & Zn	*QGFe.iari-2A, QGFe.iari-5A, QGFe.iari-7A and QGFe.iari-7B, QGZn.iari-2A, QGZn.iari-4A, QGZn.iari-5A, QGZn. iari-7A and QGZn.iari-7B*	
26.	Bubo × Turtur			
27.	Louries × Batelur			
28.	Roelfs F 2007 × Chinese Parental Line			
29.	*Hexaploid wheat (6x)* Tianong18 × Limmai6	Se		
30.	WH542 × synthetic derivative (*Triticum dicoccon* PI94624/ *Aegilops tauschii* [409]//BCN). RIL (163)			
31.	Jingdong 8 × Bainong AK58			
32.	Kachu × Zinc-Shakti			

to transfer low phytic acid (LPA)-GPC in the cultivars for enhancing Fe and Zn concentration along with grain protein content without yield penalty.

Several marker-trait associations (MTAs) for nutritional traits in wheat have been used. A total of 39 Zn MTAs were discovered in two studies, one involving 330 bread wheat genotypes and the other involving 320 genotypes from the Spring Wheat Reference Set (SWRS), with two large-effect QTL regions discovered on chromosomes 2 and 7. CIMMYT developed new biofortified varieties that are 20–40% superior in grain Zn content and agronomically on par with or better than popular South Asian wheat varieties. A GWAS study for grain Zn concentrations in 369 European wheat genotypes identified 40 MTAs on chromosomes 2A, 3A, 3B, 4A, 4D, 5A, 5B, 5D, 6D, 7A, 7B, and 7D, with the important and reliable MTAs having significant effects were localized on chromosomes 3B (723,504,241–723,611,488 bp) and 5A (462,763,758–466,582,184 bp). Reported an increase in the number of MTAs to 161 genomic regions, including recently identified candidate genes for Zn uptake and transport. A GBS study on a panel of 167 Ae. tauschii accessions revealed 5249 markers, as well as wide variability in micronutrient concentrations. Overall, 19 SNP MTAs were found on all 7 chromosomes, with positive associations found for 5 with grain Fe and 4 with Zn content. These associations were found to be associated with

genes that code for transcription factor regulators, transporters, and phytosiderophore synthesis. With improved genomic prediction accuracy and lower genotyping costs, genomic selection in wheat breeding has gained traction (Juliana et al. 2019; Charmet et al. 2020). In the coming years, the ability of genomic selection to increase genetic gain for quality traits will further transform wheat improvement methodologies. Genomic selection would aid in the rapid selection of desired plant types by using widely distributed markers, estimating the effects of all loci, and accurate prediction of genomic estimated breeding values with precise genotypic and phenotypic data (GEBV). Linear prediction models such as G-BLUP and machine learning algorithms are used to recognize complex data patterns and draw appropriate conclusions, as well as to exploit GxE interactions. Using multi-trait and multi-environmental models improves prediction accuracy and performance in selected breeding. Overall, genomic selection improves selection accuracy while decreasing time and cost for varietal development, particularly for complex characters with low heritability that are difficult to improve using traditional plant breeding methods (Heffner et al. 2009). An efficient genomic selection approach optimizes the statistical prediction model for developing GEBVs based solely on genotyping data for an un-phenotyped population using a precisely genotyped and phenotyped "training" population. This reduces the breeding cycle and allows breeders to dispense with avoidable multi-location and multi-environmental testing of genotypes.

When predicting the genotypic value of one panel based on another, prediction accuracy decreased. These findings are consistent with those of who demonstrated the high prediction accuracy in a germplasm set with high variation is used as training population as it has high genomic coverage. The use of efficient genomic selection models in conjunction with precision phenotyping on genetically variable populations will improve prediction accuracy, selection efficiency, and speed up the varietal development process. The addition of GWAS and genomic selection to MAS will undoubtedly shorten breeding cycles and improve the breeders' equation. As a result, there is enormous potential for scaling up biotechnological approaches with traditional plant breeding for varietal development.

6 Role of Nanotechnology for Nutritional Improvement of Wheat

Nanotechnology has recently emerged as one of the outstanding technologies that can be gradually applied in agriculture for crop biofortification and can largely avoid the drawbacks associated with genetic and traditional agronomic biofortification. Nanomaterials have a variety of advantageous properties, including controlled and slow discharge at target sites, significantly higher absorption capacity, and a high volume to surface ratio for effective use in the production of nanofertilizers (NFs). Because nanofertilizers are used in minute quantities, they prevent the accumulation of residual by-products of chemical fertilizers in soil and thus have a lower environmental impact. Furthermore, nanofertilizers can be generated using biosensors based on soil status and crop nutritional demand.

Furthermore, NFs have been shown to improve crop performance under various stresses by modulating carbohydrate and protein synthesis, seedling growth, nitrogen metabolism, photosynthesis, and nutrient mobilization from the rhizosphere to specific plant parts. Because of its target-bound slow delivery, even a trace amount of NFs can effectively improve micronutrient levels in grain without negatively impacting the environment. Wheat, as a major staple crop, has always been a driving force behind its use in various types of biofortification and fortification. As with other biofortification strategies, a number of greenhouse or small-scale field studies show that nanomaterials have a positive effect on wheat nutritional content. Large-scale experiments in open and enclosed areas are required to assess the benefits and drawbacks of NF-based wheat nutrient enrichment. To increase the use of NFs in large-scale wheat biofortification programs, we must first understand how different nanomaterials, their combinations, and application strategies affect target nutrient levels in different wheat genotypes. Significant attention has recently been directed toward developing appropriate methodologies for applying nanoparticles, as this significantly affects the extent of micronutrient accumulation in wheat plants.

7 Role of Genome Editing Technology in Wheat Nutritional Quality Improvement

Gene editing technology has proven to be an effective tool for modifying desired traits in a variety of crop plants, including wheat. Modifying nutritional traits in crop plants provides several health benefits, particularly in staple crops like wheat. In this section, we will discuss the current progress made in the use of gene editing technology to improve the nutritional quality of wheat. Liang et al. (2017) demonstrated direct editing of the *TaGASR7* and *TaGW2* genes by introducing the CRISPR-Cas9 ribonucleoprotein complex into immature embryos of the wheat varieties Kenong 199 and YZ814. Created heritable mutations in the *TaLpx-1, TaGW2,* and *TaMLO* genes. They also demonstrated that the seed size and thousand grain weight were significantly larger (TGW). They also discovered that TaGW2 has a negative effect on grain size in wheat. In 2019, Jouanin and colleagues investigated the possibility of simultaneously editing multiple genes in the large – and – gliadin gene families.

8 Nutritional Improvement in Wheat Quality Using Genetic Engineering

Biotechnological crop improvement interventions have proven their worth for a variety of crops, including cereals, over the last 2–3 decades. Bt cotton, maize, and golden rice are a few examples. In the case of wheat, various genes determine traits such as starch composition, nutritional profile, and final end product formulation, all of which affect final grain quality. The cloning of the NAC gene *Gpc-B1* for

grain protein content from this accession's chromosome arm 6BS encodes a transcription factor that accelerates senescence in plant vegetative parts, resulting in increased nutrient mobilization and mineral (iron and zinc) transfer to the grain (Shewry 2009). This could help in the fight against malnutrition. According to USDA World Wheat Collection research, the protein content of wheat grain can range from 7 to 22%, with 33% controlled by genomics and the rest by environmental factors, making breeding for this trait difficult (Vogel et al. 1978). This bottleneck can be circumvented by incorporating sources of variation from exotic bread wheat lines or related wild species, such as Atlas 50 and Atlas 66, which are derived from the South American line Frandoso. Both lines appeared to have multiple genes for high protein content in grain and have been extensively used in Nebraska breeding programs. Johnson et al. (1985) successfully inserted the Atlas 66 gene into the commercial variety Lancota (Johnson et al. 1985; Shewry 2009).

Biotechnological interventions have been attempted to increase grain starch while also modulating its quality (Borisjuk et al. 2019). *TaRSR1*, a wheat homolog of Rice Starch Regulator (*OsRSR1*), is a transcription factor that negatively regulates the gene expression pattern of some starch synthesis-related enzymes in wheat grains. Downregulation of *TaRSR1* resulted in a nearly 30% increase in starch content and a 20% increase in yield (Kang et al. 2013). Increased amylose in starch contributes to resistant starch (RS) in food, which can offer protection from health conditions such as diabetes, obesity, and cardiovascular diseases, many of which are chronic diseases (Borisjuk et al. 2019; Meenu and Xu 2019). A number of experiments focused on downregulation of starch branching enzymes SBEIIa and SBEIIb, leading to substantially elevated levels of amylose and resistant starch in wheat, which could benefit human health viz. obesity (Vetrani et al. 2018).

The level of free amino acids in wheat was significantly altered when the GCN2-type protein kinase gene was overexpressed. In another experiment, the pA25-TaGW2-RNAi DNA construct was implanted into the immature embryos of the wheat variety 'Shi 4185,' resulting in *TaGW2* gene suppression and increased grain weight and width. Using *Pina-D1a* and *pinb-D1b* genes, indicated the interaction of *PINA* with *PINB* to form friabilin which ultimately modulate the wheat grain texture.

Aggarwal et al. (2018) created a TaIPK1:pMCG161 RNAi construct that was then mobilized into C306 wheat genotypes. The transgenic lines reduced phytate by 28–56%, increasing the molar ratios of iron: phytic acid and zinc: phytic acid. Similarly, demonstrated a reduction in phytate of 22–34% using the TaABCC13: pMCG161 RNAi construct. Furthermore, several genes for Fe and Zn homeostasis in wheat have been identified, which are associated with four key pathways: the methionine cycle, phytosirophore biosynthesis, the transport system, and the antioxidant system. These genes could be used for gene editing or genetic engineering to increase the amount of Fe and Zn in wheat grain and its bioavailability. In general, there is ample scope to improve wheat nutritional quality components using modern approaches to improve the quality standards of commercial wheat.

9 Conclusion and Future Perspectives

Wheat grain contains various compounds which have potential nutraceutical functions. Wheat nutritional properties can be exploited in variousforms to prevent malnutrition and deadly diseases. For instance, wheat brans are the rich sources of flavonoids, phenolic acids, tocopherols, lignans, phytosterols, and carotenoids, which provide many health benefits. Wheat products contain a variety of high-value compounds, mainly bioactive compounds with significant health benefits. They can be exploited as food ingredients, supplements, additives, or extracts that are high in functional molecules and micronutrients, such as zinc, iron, and manganese, phenolic compounds, novel carbohydrates, carotenoids, biopeptides, bioactive fatty acids, amino acids, prebiotics, vitamins, and mineral elements. Bioactive compounds derived from wheat can be used as antioxidants and preservatives, reducing lipid oxidation and microbial growth. Furthermore, processing technologies to improve nutritional characteristics and sensory features also been targeted to increase the functional food value, and nutrients bioavailability, while reducing the anti-nutritional factors of cereal by-products. In the near future, more studies are necessary on the nutraceutical properties of wheat including bioactive compounds for use as nutraceuticals or as ingredients in the development of functional products. Some of these traits can be integrated in the wheat breeding activities as it is a key set of characteristics for the trading and commercialization of the grain. Grain nutritional quality should be an integral part of the breeding process and considered within the variety development process to deliver new products with better nutritional properties to consumers.

References

Aggarwal S, Kumar A, Bhati KK, Kaur G, Shukla V, Tiwari S, Pandey AK (2018) RNAi-mediated downregulation of inositol pentakisphosphate kinase (IPK1) in wheat grains decreases phytic acid levels and increases Fe and Zn accumulation. Front Plant Sci 9:259

Alotaibi F, Alharbi S, Alotaibi M, Al Mosallam M, Motawei M, Alrajhi A (2021) Wheat omics: classical breeding to new breeding technologies. Saudi J Biol Sci 28(2):1433–1444

Andersson AA, Andersson R, Piironen V, Lampi AM, Nyström L, Boros D et al (2013) Contents of dietary fibre components and their relation to associated bioactive components in whole grain wheat samples from the HEALTHGRAIN diversity screen. Food Chem 136(3–4):1243–1248

Batifoulier F, Verny MA, Chanliaud E, Remesy C, Demigné C (2006) Variability of B vitamin concentrations in wheat grain, milling fractions and bread products. Eur J Agron 25(2):163–169

Belderok B, Mesdag J, Donner DA (2000) The wheat grain. In: Belderok B, Mesdag H, Donner DA (eds) Bread-making quality of wheat. Springer, Dordrecht, pp 15–20

Bonnett DG, Rebetzke GJ, Spielmeyer W (2005) Strategies for efficient implementation of molecular markers in wheat breeding. Mol Breed 15(1):75–85

Borisjuk N, Kishchenko O, Eliby S, Schramm C, Anderson P, Jatayev S et al (2019) Genetic modification for wheat improvement: from transgenesis to genome editing. BioMed Res Int 2019:6216304

Brevis JC, Dubcovsky J (2010) Effect of chromosome region including the *Gpc-B1* locus on wheat protein and protein yield. Crop Sci 50:93–104

Brouns FJ, van Buul VJ, Shewry PR (2013) Does wheat make us fat and sick? J Cereal Sci 58(2): 209–215

Cakmak I (2008) Enrichment of cereal grains with zinc: agronomic or genetic biofortification? Plant Soil 302(1):1–17

Charmet G, Tran LG, Auzanneau J, Rincent R, Bouchet S (2020) BWGS: AR package for genomic selection and its application to a wheat breeding programme. PLoS One 15(4):e0222733

Choge JK (2020) Malnutrition: current challenges and future perspectives. In: Imran M, Imran A (eds) Malnutrition. IntechOpen, USA, pp 1–56

Cummins AG, Roberts-Thomson IC (2009) Prevalence of celiac disease in the Asia–Pacific region. J Gastroenterol Hepatol 24(8):1347–1351

Development Initiatives (2020) 2020 global nutrition report: action on equity to end malnutrition. Development Initiatives Poverty Research Ltd, Bristol. https://globalnutritionreport.org/reports/2020-global-nutrition-report/

Distelfeld A, Uauy C, Olmos S, Schlatter AR, Dubcovsky J, Fahima T (2004) Microcolinearity between a 2-cM region encompassing the grain protein content locus *Gpc-6B1* on wheat chromosome 6B and a 350-kb region on rice chromosome 2. Funct Integr Genom 4(1):59–66

Distelfeld A, Uauy C, Fahima T, Dubcovsky J (2006) Physical map of the wheat high-grain protein content gene *Gpc-B1* and development of a high-throughput molecular marker. New Phytol 169 (4):753–763

Dubcovsky J, Dvorak J (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. Science 316(5833):1862–1866

Evenson RE, Gollin D (2003a) Assessing the impact of the Green Revolution, 1960 to 2000. Science 300:758–762

Evenson RE, Gollin D (2003b) Crop variety improvement and its effect on productivity the impact of international agricultural research. CAB International, Wallingford

Feldman M (1995) Wheats. In: Smartt J, Simmonds NW (eds) Evolution of crop plants. Longman Scientific and Technical, Harlow, pp 185–192

Feldman M (2001) Origin of cultivated wheat. In: Bonjean AP, Angus WJ (eds) The world wheat book: a history of wheat breeding. Lavoisier Publishing, Paris, pp 3–56

Gonzalez-Thuillier I, Salt L, Chope G, Penson S, Skeggs P, Tosi P et al (2015) Distribution of lipids in the grain of wheat (cv. Hereward) determined by lipidomic analysis of milling and pearling fractions. J Agric Food Chem 63(49):10705–10716

Govindan V, Singh RP, Juliana P et al (2022) Mainstreaming grain zinc and iron concentrations in CIMMYT wheat germplasm. J Cereal Sci 105:103473. https://doi.org/10.1016/J.JCS.2022.103473

Gupta OP, Karkute SG (eds) (2021) Genome editing in plants: principles and applications. CRC Press, Boca Raton

Gupta PK, Langridge P, Mir RR (2010) Marker-assisted wheat breeding: present status and future possibilities. Mol Breed 26(2):145–161

Gupta OP, Deshmukh R, Kumar A, Singh SK, Sharma P, Ram S, Singh GP (2022) From gene to biomolecular networks: a review of evidences for understanding complex biological function in plants. Curr Opin Biotechnol 74:66–74

Guzmán C, Medina-Larqué AS, Velu G, González-Santoyo H, Singh RP, Huerta-Espino J et al (2014) Use of wheat genetic resources to develop biofortified wheat with enhanced grain zinc and iron concentrations and desirable processing quality. J Cereal Sci 60(3):617–622

He XY, He ZH, Ma W, Appels R, Xia XC (2009) Allelic variants of phytoene synthase 1 (Psy1) genes in Chinese and CIMMYT wheat cultivars and development of functional markers for flour colour. Mol Breed 23(4):553–563

Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. Crop Sci 49: 1–12

Henderson KN, Tye-Din JA, Reid HH, Chen Z, Borg NA, Beissbarth T et al (2007) A structural and immunological basis for the role of human leukocyte antigen DQ8 in celiac disease. Immunity 27(1):23–34

Henry RJ, Rangan P, Furtado A (2016) Functional cereals for production in new and variable climates. Curr Opin Plant Biol 30:11–18

Johnson VA, Mattern PJ, Peterson CJ, Kuhr SL (1985) Improvement of wheat protein by traditional breeding and genetic techniques. Cereal Chem 62(5):350–355

Jouanin A, Schaart JG, Boyd LA, Cockram J, Leigh FJ, Bates R et al (2019) Outlook for coeliac disease patients: towards bread wheat with hypoimmunogenic gluten by gene editing of α-and γ-gliadin gene families. BMC Plant Biol 19(1):1–16

Juliana P, Montesinos-López OA, Crossa J, Mondal S, González Pérez L, Poland J et al (2019) Integrating genomic-enabled prediction and high-throughput phenotyping in breeding for climate-resilient bread wheat. Theor Appl Genet 132(1):177–194

Kade M, Barneix AJ, Olmos S, Dubcovsky J (2005) Nitrogen uptake and remobilization in tetraploid 'Langdon'durum wheat and a recombinant substitution line with the high grain protein gene *Gpc-B1*. Plant Breed 124(4):343–349

Kaiser N, Douches D, Dhingra A, Glenn KC, Herzig PR, Stowe EC, Swarup S (2020) The role of conventional plant breeding in ensuring safe levels of naturally occurring toxins in food crops. Trends Food Sci Technol 100:51–66

Kang GZ, Xu W, Liu GQ, Peng XQ, Guo TC (2013) Comprehensive analysis of the transcription of starch synthesis genes and the transcription factor RSR1 in wheat (*Triticum aestivum*) endosperm. Genome 56(2):115–122

Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533(7603):420–424

Langridge P (2013) Wheat genomics and the ambitious targets for future wheat production. Genome 56(10):545–547

Liang Z, Chen K, Li T, Zhang Y, Wang Y, Zhao Q et al (2017) Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. Nat Commun 8(1):1–5

Meenu M, Xu B (2019) A critical review on anti-diabetic and anti-obesity effects of dietary resistant starch. Crit Rev Food Sci Nutr 59(18):3019–3031

Mohan D, Sendhil R, Gupta OP, Pandey V, Krishnappa G, Singh GP (2022) Wheat quality index: new holistic approach to identify quality superior genotypes. Cereal Res Commun 50:1–11

Oury FX, Leenhardt F, Remesy C, Chanliaud E, Duperrier B, Balfourier F, Charmet G (2006) Genetic variability and stability of grain magnesium, zinc and iron concentrations in bread wheat. Eur J Agron 25(2):177–185

Peleg Z, Saranga Y, Yazici A, Fahima T, Ozturk L, Cakmak I (2008) Grain zinc, iron and protein concentrations and zinc-efficiency in wild emmer wheat under contrasting irrigation regimes. Plant Soil 306(1):57–67

Peleg Z, Cakmak I, Ozturk L, Yazici A, Jun Y, Budak H et al (2009) Quantitative trait loci conferring grain mineral nutrient concentrations in durum wheat × wild emmer wheat RIL population. Theor Appl Genet 119(2):353–369

Piironen V, Edelmann M, Kariluoto S, Bedo Z (2008) Folate in wheat genotypes in the HEALTHGRAIN diversity screen. J Agric Food Chem 56(21):9726–9731

Poole N, Donovan J, Erenstein O (2021) Agri-nutrition research: revisiting the contribution of maize and wheat to human nutrition and health. Food Policy 100:101976

Pu ZE, Ma YU, He QY, Chen GY, Wang JR, Liu YX et al (2014) Quantitative trait loci associated with micronutrient concentrations in two recombinant inbred wheat lines. J Integr Agric 13(11): 2322–2329

Ram S, Govindan V (2020) Improving wheat nutritional quality through biofortification. In: Igrejas G, Ikeda TM, Guzmán C (eds) Wheat quality for improving processing and human health. Springer-Nature, Switzerland AG, pp 205–222

Shewry PR (2009) Wheat. J Exp Bot 60(6):1537–1553

Shewry PR, Van Schaik F, Ravel C, Charmet G, Rakszegi M, Bedo Z, Ward JL (2011) Genotype and environment effects on the contents of vitamins B1, B2, B3, and B6 in wheat grain. J Agric Food Chem 59(19):10564–10571

Šramková Z, Gregová E, Šturdík E (2009) Chemical composition and nutritional quality of wheat grain. Acta Chim Slov 2(1):115–138

Srinivasa J, Arun B, Mishra VK, Singh GP, Velu G, Babu R et al (2014) Zinc and iron concentration QTL mapped in a Triticum spelta × T. Aestivum cross. Theor Appl Genet 127:1643

Stone B, Morell MK (2009) Carbohydrates. In: Khan K, Shewry PR (eds) Wheat chemistry and technology, 4th edn. American Association of Cereal Chemists, St Paul, pp 299–362

Tekin M, Cengiz MF, Abbasov M, Aksoy A, Canci H, Akar T (2018) Comparison of some mineral nutrients and vitamins in advanced hulled wheat lines. Cereal Chem 95(3):436–444

Uauy C (2017) Plant genomics: unlocking the genome of wheat's progenitor. Curr Biol 27(20): R1122–R1124

Velu G, Tutus Y, Gomez-Becerra HF, Hao Y, Demir L, Kara R, Crespo-Herrera LA et al (2017) QTL mapping for grain zinc and iron concentrations and zinc efficiency in a tetraploid and hexaploid wheat mapping populations. Plant Soil 411:81–99

Vetrani C, Sestili F, Vitale M, Botticella E, Giacco R, Griffo E et al (2018) Metabolic response to amylose-rich wheat-based rusks in overweight individuals. Eur J Clin Nutr 72(6):904–912

Vogel KP, Johnson VA, Mattern PJ (1978) Protein and lysine contents of endosperm and bran of the parents and progenies of crosses of common wheat. Crop Sci 18(5):751–754

Weickert MO, Pfeiffer AF (2008) Metabolic effects of dietary fiber consumption and prevention of diabetes. J Nutr 138(3):439–442

Xu Y, An D, Liu D, Zhang A, Xu H, Li B (2012) Molecular mapping of QTLs for grain zinc, iron and protein concentration of wheat across two environments. Field Crops Res 138:57–62

Maize Nutraceutomics: Genomics, Biotechnology, and Nanotechnology

Deepti B. Sagare, Prashant Shetti, Shrikant Yankanchi, Sai Rekha Kadirimangalam, Rachana Baguda, Fan Xingming, Jun Fan, Shweta Singh, Rani Asaram Jadhav, M. A. Ashrutha, and Kumari Aditi

Contents

D. B. Sagare (✉)
Bayer Crop Science Ltd, Hyderabad, India

P. Shetti
Banashankari Agro Agency, Nipani, India

S. Yankanchi
Indira Gandhi Krishi Vishwavidyalaya, Raipur, India

S. R. Kadirimangalam
International Crop Research Institute for the Semi-Arid Tropics, Hyderabad, India

R. Baguda
Professor Jayashankar Telangana State Agricultural University, Hyderabad, India

F. Xingming · J. Fan
Yunnan Academy of Agricultural Sciences, Kunming, China

S. Singh
Indian Institute of Sugarcane Research, Lucknow, India

R. A. Jadhav
College of Agriculture, Nagpur, India

M. A. Ashrutha
Hytech seeds India Pvt. Ltd, Hyderabad, India

K. Aditi
Kansas State University, Manhattan, KS, USA

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_3

Abstract

Maize (*Zea mays*) is known as the queen of cereals because of its higher grain yield potential, wider adaptability, genetic diversity, use as food, feed, and industrial use among cereals. The growing population is a threat to nutritional security, and demand for a cost-effective and promising strategy for the food system. Every year around two billion of the world's population is prone to malnutrition caused by essential micronutrients, proteins, and vitamins. Biofortification is one of the most promising approaches to enhance the nutritional quality of maize grains and reduce the risk of hidden hunger globally. Maize possesses several naturally existing mutants for nutritional quality traits and is considered as a model crop for biofortification. Effective utilization of recent advances in genomic, molecular tools, crop improvement techniques, machine learning, and artificial intelligence can pave the way in developing nutritionally rich maize. This chapter provides insight on the importance of maize in global nutritional food security, approaches for biofortification, genetic resources, and genetic diversity for nutritional quality traits, map-based gene cloning, trait mapping and finding major quantitative trait loci (QTLs), conventional and genomic-assisted breeding strategies for enhancing corn nutritional

quality. Further this chapter emphasizes the genetic engineering approach, novel genome editing techniques, nanotechnology, and available bioinformatic databases to carry out omics research for designing nutritional rich maize.

Keywords

Maize · Biofortification · Genomics · QTLs/genes · Molecular breeding · Maize genome database · Genetic engineering

1 Introduction

Corn, also called maize (*Zea mays* L.), is most prominent and versatile among cereals due to its adaptability, diversity, and use as food source for humans and animals, and wider industrial use. The literal meaning of maize is "that which sustains life." Corn is the second most widely grown crop globally in temperate, tropics, and subtropical regions. There are different types of corn such as, field corn (sweet corn, baby corn, popcorn) and specialty corn (quality protein maize (QPM), high-oil maize, waxy maize etc). In 2021, a total of 1.2 billion metric tons of corn was produced globally and the United States was the largest producer of corn followed by China and Brazil (Statista.com). Globally, 61% of corn is used as feed, 17% as food and 22% for industrial (starch and biofuel) use, and therefore corn is considered as a driver for agricultural economy. Corn consists of 72% starch, 10% protein, 4% lipid, and various vitamins except for vitamin B12. Corn germ contains 80% of mineral content whereas, less than 1% resides in the endosperm. On a dry basis, maize provides approximately1400 Kcal/100 gm of energy to perform various physiological activities. Maize is a good source of fat, dietary fibers, essential amino acids, proteins, carbohydrates, starch, b-complex vitamins (histamine, folic acid, riboflavin, niacin, pantothenic acid), fat soluble vitamins (A, D, E, K), minerals and trace elements, organic acids, polyphenols, phytosterols (Table 1).

Growing population is a threat to nutritional security, and there is a stealthy form of hunger called "micronutrient malnutrition" or "hidden hunger." Every year around two billion of the world's population is prone to malnutrition caused by key micronutrients Fe and Zn, and, provitamin A. Micronutrients play an important role to perform functions of the human body and inadequate consumption of micronutrients causes criticalities and several effects on biological A deficiency. Deficiency of protein causes marasmus, and kwashiorkor, fiber deficiency leads to diverticulitis, and constipation, whereas iron and vitamin A deficiency causes anemia and night blindness, respectively. Calf muscle pain and heart muscle weakening happen due to thiamine deficiency. Pyridoxine and niacin deficiency causes angular stomatitis and diarrhea, dermatitis, and dementia, respectively. The megaloblastic anemia is caused by folic acid deficiency, and, antioxidants deficiency decreases immunity. Thus, to minimize risk of malnutrition, foods rich in micronutrient content need to be included in the human diet. Fruits and vegetables are good sources of micronutrients. In developing countries where most of the population relies on staple cereal-based food diets either cannot have access or

Table 1 Nutrient content of maize

Nutrients	Nutritive value (as per 100 gm of edible portion)
Protein (g)	8.80 ± 0.49
Total fiber (g)	12.24 ± 0.93
Carbohydrates (g)	64.77 ± 1.58
Lysine (g)	2.64 ± 0.18
Tryptophan (g)	0.57 ± 0.12
Total starch (g)	59.35 ± 0.83
Thiamine (B1) (mg)	0.35 ± 0.039
Riboflavin (B2) (mg)	0.14 ± 0.014
Niacin (B3) (mg)	2.10 ± 0.09
Panthothenic acid (B5) (mg)	0.27 ± 0.02
Total B6 (mg)	0.28 ± 0.023
Biotin (B7) (mg)	0.70 ± 0.06
Total folates (B9) (mg)	39.42 ± 3.13
β-carotene (μg)	186 ± 19.2
B- cryptoxanthin (μg)	110 ± 10.1
Total carotenoids (μg)	893 ± 154
α-tocopherol, vitamin-E (mg)	0.36 ± 0.03
Phylloquinones, Vitamin-K (μg)	2.50 ± 0.76
Oleic acid (C18:1) (mg)	700 ± 17.9
Linolenic acid (C18:2) (mg)	1565 ± 18.2
Total saturated fatty acids (TSFA) (mg)	413 ± 5.6
Total monounsaturated fatty acids (TMUFA) (mg)	706 ± 17.4
Total polyunsaturated fatty acids (TPUFA) (mg)	1606 ± 18.5
Calcium (Ca) (mg)	8.91 ± 0.61
Iron (Fe) (mg)	2.49 ± 0.32
Total oxalate (mg)	15.26 ± 1.78
Total polyphenols (mg)	32.92 ± 3.85

affordability of fruits and vegetables. Therefore, for nutrient security of cereals various approaches need to be adopted such as, supplementation, diversification, and fortification. Biofortification is the enhancement of micronutrients in the edible part of the crop and is a most promising approach to alleviate malnutrition. As there are several naturally existing mutant alleles for nutritionally important genes have been found in corn, it is considered a model crop for biofortification (Sagare et al. 2018).

2 Methods of Biofortification

Improvement of nutritional quality of a crop should be done by elevating the micronutrient content of the edible portion without compromising yield and yield attributing characters. For enhancing nutritional quality of food grains, three

different biofortification methods are used, agronomic biofortification, transgenic biofortification and conventional biofortification. In conventional biofortification, the high yielding and popular varieties are a cross breed with nutritionally rich varieties, and then backcrossed to develop staple crops with higher nutrient content. In the genetic biofortification approach, a foreign gene (bacterial/plant) is introduced or manipulated genomic regions controlling nutritional related genes, to enhance micronutrient content in edible parts of the plant. In agronomic biofortification, application of micronutrient fertilizers, soil conditioning, or seed amendment is conducted to increase micronutrient content of dietary field crops. Soil amendment manages pH of the soil and makes micronutrients available in the root zone for plants to uptake. Soil amendment is done by application of lime, biochar, biosolids, elemental sulfur, gypsum, plant residues, animal manure, organic manure etc. Biofortification can also be done through nanotechnology, nanoparticles are incorporated into micronutrient fertilizers and exploited for their role in nutritional and yield enhancement. In corn, foliar application of zinc hydroxide nitrate (Zn5 (OH)8 $(NO_3)_2 \cdot 2H_2O$) led to accumulation of Zn content in endosperm (Ivanov et al. 2019). The conventional approaches are time consuming and are unable to fortify multiple micronutrients. In the recent past the genomic advances led to identifying various nutrition related genes and novel approaches such as, trait mapping, molecular-genomic breeding, genomic selection can accelerate breeding of staple food crops to enhance biofortification.

3 Genetic Resources for Nutritional Quality Improvement

Maize belongs to the family Poaceae and tribe Maydeae/Andropogonodae that comprises seven genera, viz. *Coix* (2n = 10 or 20), *Sclerachne* (2n = 20), *Trilobachne* (2n = 20), *Chionachne* (2n = 20) and *Polytoca* (2n = 20) (Old World groups) and *Zea* and *Tripsacum* (New World groups). Two new world group genera, *Zea* and *Tripsacum* forms the genepool of maize, and comprises subtribe Tripsacinae, tribe Andropogoneae, and subfamily Panicoideae of family Poaceae (Barker et al. 2001). A gene pool categorization of cultivated crops was proposed based on the feasibility of gene flow/transfer from those species to cultivated crop species. The categories are, primary, secondary, and the tertiary gene pools. Biological species that have no gene exchange barrier fall under the primary gene pool that comprises wild and progenitors of cultivated crop species. Secondary gene pool comprises distantly related crop species and has crossability issues. Outer limits of potential genetic resources are considered as tertiary gene pools. *Zea mays* (ssp. *mays*) represents the primary genepool, other taxa in the genus *Zea* (*teosintes*) represents secondary genepool (subspecies *parviglumis, mexicana, huehuetenangensis*, and other zea species *luxurians, nicaraguensis, diploperennis, perennis*), and all the species in the genus *Tripsacum,* sister genus to *Zea* (new world perennial, polyploid grasses) represents tertiary genepool. About 9000 years ago maize (*Zea mays* ssp. *mays*) was domesticated from wild species ancestor, teosinte (*Zea mays* ssp. *parviglumis*).

The wild relative serves as, great source genetic diversity and valuable resources of economically important genes viz., biotic and abiotic stress resistance, yield related genes and genes comprising higher nutrient content, and can be used in crop improvement. Wild relatives are a source of alleles for important traits such as insect resistance and gray leaf spot (GLS) resistance from *Z. mays* ssp. *mexicana* (Lennon et al. 2016), waterlogging resistance from *Z. diploperennis* and *Z. nicaraguensis*, resistance to the parasitic weed Striga from *Z. diploperennis* (Amusan et al. 2008), higher yield from *Z. mays* ssp. *Parviglumis* (Wang et al. 2008). Several nutritional quality-related mutant alleles were identified from cultivated wild relatives and landraces viz., *waxy1* for high amylopectin, *opaque2* for high quality protein, *Y1* for high carotenoid content (Amusan et al. 2008). Maize lines rich in methionine, tryptophan and lysine were obtained by allelic introgression from *Z. mays* subsp. *mexicana* (Wang et al. 2008). Teosinte possesses several functional variations for nutritional quality traits, protein content, provitamin A carotenoids, Starch and oil content (Harjes et al. 2008; Karn et al. 2017), therefore, it could serve as a potential donor for maize biofortification.

4 Genetic Diversity Analysis for Nutritional Quality-Related Genes

Genetic diversity changes over time and space can be defined as the range of genetic characteristics presented within a crop or species (Swarup et al. 2021). Genetic diversity analysis plays a fundamental role in breeding, plant genetics, biological evolution, and biodiversity conservation. Understanding maize genetic difference and its distribution is particularly important for selecting diverse parental maize materials to improve maize quality. The analyses of genetic diversity can be classified as two categories: one is the phenotype-based diversity analysis which focuses on the observation of morphological traits such as height and color, while the other one is molecular-based analysis which focuses on the examination of DNA variations. The maize genetic diversity has strong relationships with its wild relatives, so conserving and utilizing wild relatives could help improve maize breeding. In addition, the distribution of maize genetic diversity shows obvious geographical characteristics (Mir et al. 2013).

4.1 Morpho-Pheno-Biochemical Traits-Based Diversity Analysis

Morpho-phenotype-based genetic diversity analysis refers to the use of analysis of phenotypic differences among plants to further infer genetic diversity, whereas, biochemical-diversity reveals the nutrient content of the seeds. It is widely used in new cultivar breeding, evaluation of genetic response to environment, and crop management such as biological pest control (Swarup et al. 2021). Diversity analysis based on kernel composition, and zein profiles in a diverse set of modern inbred lines, teosinte accessions, landraces, intermediate between inbred and teosinte, revealed

in-spite of having smaller seed size teosinte possesses twice the protein content of landraces and inbred lines. The diversity analysis study based on morpho-phenotypic (plant height, kernel length, grain weight, kernel rows per cob, cob diameter etc.), biochemical/nutritional content (tryptophan) and kernel color (orange, white, yellow) traits in 1348 accessions of 13 maize populations belonging to stiff and non-stiff stalk heterotic groups, revealed variable relationships between and within total diversity among morpho-pheno-biochemical traits (Jaradat and Goldstein 2013). The diversity analysis study conducted in 51 accessions of maize landraces from north west Himalayas of India based on agro-morphological (grain yield per plant, plant height, ear height, kernel rows, kernels per row etc.) and nutritional quality (tryptophan content) traits, revealed significant differences among the accessions (Kumar et al. 2015). A total of 1279 accessions conserved in the Indonesian Agency for Agricultural Research and Development-Indonesian Center for Agricultural and Biotechnology and Genetic Resources Research and Development Gene Bank (IAARD-ICABIOGRAD Gene Bank) were assessed for their kernel characteristics (morphological- color, size, and shape) and a moderate-high genetic diversity among accessions was observed (Risliawati et al. 2022). Morpho-phenotype-based diversity analysis can detect unique traits or genotypes that can be used in breeding programs. A weakness of phenotype-based diversity analysis is that homozygotes and heterozygotes are not distinguishable. Also, as phenotype is affected not only by genes but also by environment, morpho-phenotype-based diversity analysis may need to be further validated by other methods such as molecular marker assistant analysis. Whereas, biochemical or nutritional content based genetic diversity is very much important to find out nutritionally rich germplasm and their subsequent utilization in breeding nutritionally rich crops.

4.2 Molecular Marker-Based Diversity Analysis

A molecular marker is a DNA sequence that reveals mutation or variation and can be located and identified in the genome. A number of metrics are used to evaluate genetic diversity such as the number of alleles, genetic differentiation index, observed or expected heterozygosity, percentage of polymorphic loci, genetic distance, and polymorphism information content. Moreover, hierarchical analysis of molecular variance and principal coordinate analysis are also widely used in genetic diversity analysis (Wu et al. 2021). Several studies have been taken up for molecular marker based genetic diversity for nutritional quality traits. Molecular diversity in 20 inbred lines with varied Kernel tryptophan, Lysine, provitamin A, Fe and Zn content, using 25 simple sequence repeat (SSR) markers, revealed high levels of polymorphisms, and the nutritionally contrasting could and genetically diverse lines can serve as potential sources for hybrid development program as well as for further studies on quantitative trait locus (QTL) analysis of kernel micronutrient traits (Jaiswal et al. 2019). A biochemical (α-tocopherol or vitamin E content) and molecular analysis (SSRs-based diversity) of set of 24 inbreds possessing favorable alleles of **γ-***tocopherol methyltransferase* (*ZmVTE4*) identified set of potential cross

combinations for developing high yielding vitamin E rich hybrids, and to map additional genes affecting accumulation of α-tocopherol (Das et al. 2019).

5 Classical Genetics and Traditional Breeding for Nutritionally Rich Maize

Classical genetics studies the mechanism of how certain characteristics are passed from parents to offspring without knowing the molecular details, and traditional breeding is based on the variations of observable traits. Breeders look for desired traits and use traditional methods such as mass selection, recurrent selection, hybridization, and synthetic variety development to create new cultivars.

5.1 Genetics of Nutritional Quality-Related Genes

Starch accounts for about 70% of maize kernel weight and is critical for grain yield and quality. In maize kernel sucrose is firstly converted to glucose and then further to starch, and starch is composed of amylopectin and amylose. Genetics studies reveal that starch production in maize is regulated by a wide range of genes such as *sh1, sh2, bt2, wx1, ae1*, and *su1*. These genes act in different pathways. For example, *sh1* helps the formation of glucose, *sh2* and *bt2* help the conversion from UDP-glucose to ADP-glucose, *ae1* and *su1* aid to form amylopectin, and *wx1* helps form amylose. About 10% of maize kernel weight is protein, made up of amino acids, and some of these are essential for human health. Proteins can be classified to two types: zein protein and non-zein protein. Zein protein accounts for about 70% of total protein in the maize kernel and is affected by abundant gene mutation. Some genes such as *De-B30*, *mc*, *fl2*, and *fl4* encode zein protein, while others such as *fl1* and *o1* encode non-zein protein. Moreover, genes can also regulate endosperm metabolism enzymes (*o5*, *o6*) or transcription process (i.e., *o2*, *fl3*) and ultimately affect protein (Khan et al. 2019). Probably, the most widely studied gene is *o2* which can be used to breed quality protein maize, as *o2* can reduce alpha zein accumulation but increase lysine and tryptophan contents that are important for human health (Holding 2014). Oil is also a major component in maize seed and is rich in polyunsaturated fatty acids and energy. The quality of maize oil is determined by not only the oil content but also the composition ratios among fatty acids. Genetic studies have identified several main genes that control oil-related traits. For example, *DGAT1-2* encodes diacylglycerol acyltransferase which catalyzes the final step of oil synthesis, *fad2* encodes oleate desaturase that affects oleic acid composition, and *ZmWRI1* encodes a transcription factor that can increase oil content.

5.2 Breeding Objectives

In maize breeding, selection is the main method used to eliminate or exhibit target characteristics, and it could be either natural selection or artificial selection or

selective breeding. Selective breeding aims to produce offspring that have desirable traits which can be passed to future generations via genes. The selection of multiple traits simultaneously is conducted by developing a selection index to maximize genetic gains. From the selection objectives point of view, the selection could also be classified as positive selection and negative selection, which can be translated as keeping the favorable traits and eliminating detrimental traits, respectively. Breeding varieties with enhanced nutritional qualities such as high vitamin A content, high-quality protein content, and high zinc content, while having stable yields under disease damage and/or environmental stress such as heat and drought, may be an ideal breeding objective to bred nutrition-rich maize. However, breeding objectives should also consider temporal and spatial conditions to address local specific challenges.

5.3 Classical Breeding Achievements

In the past decades, breeders have bred many nutritious maize varieties such as high-amylose maize, high-oil maize, high-vitamin maize, sweet maize, waxy maize, and quality protein maize, using traditional methods. High-amylose maize is widely used in the food industry as the high-amylose starch in maize produces opaque gels that are good materials for confectionery and thickener. The high content of amylose in maize starch is mainly controlled by a recessive gene *ae*. Through hybridization, random mating, and mass selection, breeders managed to develop the SU amylose-extender maize cultivar with apparent amylose content ranging from 53.3% to 69% with an average of 61.7%. However, the yield of the amylose-extender maize was low (3–4 tons per ha) compared with normal maize hybrids. The Germplasm Enhancement of Maize project of the United States Department of Agriculture reported that the apparent amylose contents of their *ae*-inbred lines (H99ae, OH43ae, B89ae, and B84ae) were from 61.7% to 67.7% (Li et al. 2008). Another high-amylose maize germplasm was reported by Campbell et al. (2007) whose research showed that the apparent amylose starch content of GEMS-0067 inbred line was at least 70% and had good potential for future high-amylose maize breeding. Quality Protein Maize (QPM) is another achievement. QPM contains two enhanced amino acids: lysine and tryptophan which are essential for human and animal growth and development. The breeding of QPM started in the 1960s since scientists discovered some gene mutants such as *o2* and *fl2* could improve the contents of lysine and tryptophan (Prasanna et al. 2001). The International Maize and Wheat Improvement Center (CIMMYT) successfully developed a group of *o2* based QPM gene pools (Table 2), using backcross recurrent selection. Later, by using CIMMYT's germplasm, researchers around the world bred many local QPM hybrids or open-pollinated varieties to adapt local ecosystems such as HQ INTA-993 in Nicaragua, ICA in Colombia, INIA in Peru, FONAIAP in Venezuela, BR473 and Assum Preto in Brazil, H-553C and VS-538C in Mexico, Susuma in Mozambique, Obatampa in Bein and Guinea, Obangaina in Uganda, GH-132-28 in Ghana, QS-7705 in South Africa, BHQPY545 and AMH760Q in Ethiopia, Zhongdan

Table 2 the QPM gene pools and their corresponding protein contents. (Source: Prasanna et al. 2001)

Gene pool number	Ecological adaptation	Maturity	Lysine content (%)	Tryptophan content (%)	Protein content (%)
Pool 15	Tropical	Early	4.2	0.94	9.1
Pool 17	Tropical	Early	4.5	1.04	8.9
Pool 18	Tropical	Early	4.0	0.93	9.9
Pool 23	Tropical	Late	3.8	1.03	9.1
Pool 24	Tropical	Late	3.8	0.92	9.4
Pool 25	Tropical	Late	4.0	0.94	9.8
Pool 26	Tropical	Late	4.1	0.90	9.5
Pool 27	Subtropical	Early	4.2	1.05	9.5
Pool 29	Subtropical	Early	4.3	1.06	9.2
Pool 31	Subtropical	Medium	4.1	0.96	10.2
Pool 32	Subtropical	Medium	4.2	1.04	8.9
Pool 33	Subtropical	Medium	–	1.05	9.3
Pool 34	Subtropical	Medium	4.1	1.10	9.1

9409 and Yunrui1 in China, Shaktiman in India, and HQ-2000 in Vietnam (Prasanna et al. 2001; Teklewold et al. 2015). Besides QPM, other amino acids enriched maize varieties are also reported. For example, high-methionine inbred lines derived from A632, B73 and Mo17 are registered in the United States. Breeding high-oil maize is challenging because oil content and grain yield has a negative relationship. With decades of effort, the University of Illinois selected two high-oil synthetics: Alexho Elite and UHO with oil contents accounting for 21.2% and 15%, respectively. By using this two high-oil germplasm, Lambert et al. (1998) managed to breed high-oil maize hybrids (oil content ranged from 7% to 9%) without grain yields decreasing. However, if the oil contents were increased to a range of 10–15%, the grain yield decreased. By incorporating CML171 (a tropical high-oil inbred line) and YML107 (an inbred line derived from Suwan1), the high-oil hybrid Yunrui8 is developed with oil content of 9% and yield of 9.8 ton per hectare. From 2005 to 2015, Yunrui8 has cultivated over 0.6 million hectares cumulatively in Yunnan and other Chinese provinces.

5.4 Limitations of Traditional Breeding and Rationale for Molecular Breeding

The general process of traditional breeding can be simply described as; setting breeding objectives, germplasm creation, crosses and selections, field trials and evaluations, and finally the release of new cultivars. If desirable traits do not appear, breeders need to create more mutations or introduce other germplasm and repeat the above process to obtain desirable varieties. Although traditional breeding has had many successes and contributed greatly to global food security, its limitations cannot

be ignored such as (1) longer time required to breed pure lines that have stable and inheritable traits, (2) the offspring of hybrid F_1 have inferior performance compared with F_1 hybrids, causing farmers to buy F_1 seeds annually, (3) the artificial selection narrows the gene pool, (4) the interested traits are not precisely transferred when plants are crossed, and (5) crosses can only be done between two maize plants that can be sexually mated, it is thus impossible to add useful traits or genes from other species to maize.

Given the limitations of traditional breeding, molecular breeding can overcome some of these disadvantages. By using molecular tools to track within-genome variations, molecular breeding can monitor the recombination of specific genes or marker profiles in the breeding process.

6 Mapping Grain Quality Genes and Quantitative Trait Loci (QTLs)

Maize being an economically important cereal, attention has been paid to enhance maize grain quality (Protein, oil content, Fe, Zn, provitamin A, starch). Genetic studies have revealed that nutritional quality of maize kernels is possessed by quantitative traits. The recent advances in molecular and genomic research, and high throughput phenotypic techniques have accelerated genetic dissection of nutritional quality in maize by QTL mapping. Several nutritional quality-related quantitative trait loci (QTLs) in maize have been mapped using various DNA based markers viz., amplified fragment length polymorphisms (AFLPs), restriction fragment length polymorphisms (RFLPs), simple sequence repeats (SSRs), and single nucleotide polymorphisms (SNPs), and mapping populations F_2, $F_{2:3}$, DH (doubled haploids), BC (backcross), RILs (recombinant inbred lines), NILs (near isogenic line).

6.1 QTLs for Quality Protein Maize (QPM)

Maize kernels lack essential amino acids, lysine and tryptophan, and discovery of a spontaneous mutant with opaque and soft grains called *opaque2* (*o2*) during 1940s led to the development of QPM. The homozygous recessive *o2* allele was found to enhance lysine and tryptophan content (+69%) in maize grain endosperm. Several other mutants having different effects on lysine content in maize grain were identified viz., *floury2* (*fl2*), *opaque7* (*o7*), *opaque6* (*o6*), *floury3* (*fl3*), *mucronate* (*Mc*), *defective endosperm* (*De-B30*), *opaque7749*, *opaque7455* (*o11*), and *opaque16*. In the early 90s a total of 22 loci associated with protein and two major QTLs involved in *o2* modifications were reported. Recessive mutant *opaque 16* on chromosome 8 from Robertson's Mutator stock, and two SSRs umc1141 and umc1149 linked with the *o16* mutant were reported in late 90s. Research at CIMMYT identified markers representing grain hardness (endosperm modifiers) and high amino acid contents (amino acid modifiers) in elite QPM × QPM crosses. Babu and Prasanna 2014 performed BSA (bulk segregant analysis) in phenotypically contrasting progenies of

seven QPM × QPM populations using genome-wide SNP and found genomic regions associated with grain hardness and high tryptophan contents. Babu et al. (2015) mapped a total of five significant QTLs on chromosomes 5, 7, and 9 for tryptophan content in the $F_{2:3}$ population from a cross between two isogenic QPM lines VQL2 and VQL8.

6.2 QTLs for Oil Content

Maize oil is high in polyunsaturated fatty acids and low in linolenic acid, making it a desirable vegetable oil (Lambert 2001). Improving the quantity and quality of maize kernel oil content is an important target for breeding. Number of QTL mapping studies for oil content in maize kernels have been conducted. More than 120 QTLs for oil content using different mapping populations and markers have been reported till date. A major QTL located on chromosome 6 was identified from different populations (F_2, BC, RILs) and was consequently cloned. The oil content is influenced by a few QTLs with large effects, and epistasis is also crucial in the genetic basis of maize kernel oil content.

6.3 QTLs for Starch Content

Starch is the most important component of maize kernels, accounting for 70% of the kernel weight. It is used as a raw material in industries and in the production of other products like high fructose corn syrup, polymer-based fibers, and fuel ethanol. As a result, manipulating starch quality and quantity in maize kernels is a critical goal in maize breeding. QTLs for maize starch content are reported on all ten chromosomes. More than 50 QTLs for starch content has been reported so far. Recently, a major QTL *Qsta9.1* for starch content in a 1.7 Mb interval on chromosome 9 in the RIL population has been reported by Lin et al. (2019).

6.4 QTLs for Fe, Zn, and Provitamin A Content

Micronutrient malnutrition, particularly zinc (Zn), iron (Fe), and vitamin-A deficiency in diets, has sparked global concern. The genetic dissection of these traits in major cereal grains is a prerequisite for a biofortified breeding program. A few QTL mappings have also been conducted on micronutrient content in maize. For Fe and Zn content several QTLs (>50) and metaQTLs were reported, in different populations viz., RIL population, F_4 population, $F_{2:3}$ populations. The two informative genomic regions, bins 2.07 and 2.08 are associated with higher Zn and Fe content QTLs. The dissimilar number and position of QTLs for Fe and Zn noticed in different studies indicated their complex nature and it also depends on genetic material and environment.

Yellow kernel maize is grown and consumed globally but the provitamin A content is much less (~2ug/g), and efforts were taken up to identify QTLs/genes associated with higher provitamin A content of maize kernel. So far >120 QTLs are reported to be associated with provitaminA content in corn kernel, and major QTLs are co-localized with *zeta-carotene desaturase* (*ZDS*), *y1*, *y9, phytoene synthase, carotene dioxygenase* genes. The carotenoid biosynthesis pathway in maize is well studied, the significant allelic variations in two key genes of the pathway i.e., *lycopene-Ɛ-cyclase* and *β-carotene hydroxylase 1* controls accumulation of provitaminA carotenoids in maize endosperm (Harjes et al. 2008).

7 Association Mapping for Quality Traits

A genome-wide association study (GWAS) is a method used in genetic research to correlate specific genetic differences with specific traits. Due to the large coverage of historical recombination events prominent to the rapid decay of linkage disequilibrium (LD), GWAS analysis is the best approach to efficiently fine map QTL. Maize is the most suitable crop to perform GWAS due to abundant genetic variability, distinct subpopulations, plenty of SNP information, and rapid decay of LD. GWAS revealed a loci *Zmfad2* responsible for the changes in oleic acid content. Similarly, a strong candidate gene *DGAT1–2* controlling oil content on chromosome 6 was identified through GWAS. Genome-wide SNP scanning of contrasting lines for tryptophan content revealed many polymorphic regions, most notably 2.07, 5.03, and 10.03, which are associated with grain modifications and tryptophan contents. Cook et al. (2012) conducted Joint-linkage mapping (JLM) and GWAS in a nested association mapping population, and identified 21–26 QTLs through JLM for kernel oil, starch, and protein content; whereas from GWAS a gene *Acyl-CoA: diacylglycerol acyltransferase1–2*, controlling oil composition and quantity. Li et al. (2013) revealed 74 loci associated with kernel oil and folic acid composition in GWAS study using 1.03 million SNPs, further these loci were validated by expression QTL mapping and co-expression analysis. A coding region of the carotenoid biosynthetic genes *zep1* and *lut1*, as well as previously associated *lcyE* and *crtRB1* genes were identified by Owens et al. (2014) in a GWAS across 281 maize inbreds.

A *CrtRB1* gene on chromosome 10 related to β-carotene was identified from GWAS using CIMMYTs CAM panel of 380 diverse tropical and subtropical inbred lines (Suwarno et al. 2015). GWAS carried out by Liu et al. (2016) with a set of 263 maize inbred lines genotyped with the SNP50 BeadChip maize array, identified four QTLs and four genes within the 100-kb intervals and 77 candidate genes associated with starch synthesis; they further reported, *Glucose-1-phosphate adenyltransferase* (APS1; *GRMZM2G163437*) as a key controller of kernel starch content. GWAS using three RIL populations conducted by Deng et al. (2017) revealed 247 and 281 significant loci associated with grain protein controlling genes in two different environments, and reported that the *O2 (GRMZM2G015534)* gene is located approximately 98 Kb downstream of the lead SNP on chromosome

7 and regulates endosperm storage protein genes such as 22 kD and zein. A significant association between several SNPs on chromosome 9 and amylose content was revealed in a GWAS analysis (Li et al. 2018). Association mapping using high-density SNPs in 923 inbred maize lines was evaluated for Fe and Zn by Hindu et al. (2018), and a total of 20 and 26 SNPs were found to be significantly associated with Kernel Zn and Fe concentrations.

The carotenoid biosynthesis genes *crtRB1, lcyE,* and *ZEP1* were discovered in a GWAS performed using 130 diverse panels of yellow maize tropical inbred lines (Azmach et al. 2018). Furthermore, several transcription factors viz., RING zinc finger domain, and HLH DNA-binding domain superfamily proteins involved in the regulation of carotenoid biosynthesis were reported. A total of 49 SNPs significantly associated with five-grain quality traits, and three candidate genes for protein content, three for oil content, and three for starch content were identified using 83,057 SNPs in a GWAS with 248 diverse inbred lines (Zheng et al. 2021).

8 Marker-Assisted Breeding

Marker-assisted breeding (MAB) selects plants for inclusion in the breeding program at an early stage of development by using molecular markers associated with desirable traits. MAB uses molecular markers, molecular biotechnology, genomics and linkage mapping in conjunction with genotypic analyses to improve plant traits. Marker-assisted selection (MAS), marker-assisted backcrossing (MABC), marker-assisted recurrent selection (MARS), genomic selection (GS), and genome-wide selection (GWS) are all synonyms for this term. MAB has been used successfully to improve the genetics of various nutritional traits in maize. The important QTLs/genes to be utilized in molecular or genomic breeding, advanced tools and techniques to develop nutritional rich maize are depicted in Fig. 1.

8.1 MAB for Quality Protein Maize

The initiation of MAS for QPM development was accomplished with *o2*-specific SSR (Gupta et al. 2009; Shetti et al. 2020). Through MABB, the Institute of Crop Science and the Chinese Academy of Agricultural Sciences (CAAS) established several diverse QPM lines with different genetic backgrounds (Tian et al. 2004). The introgression of *o2* allele into normal maize inbred lines has resulted in a 41% and 30% increase in tryptophan and lysine contents, when compared to non-QPM hybrids. The MAS-derived QPM version of "Vivek Hybrid 9" was designated as "Vivek QPM 9" possessing significantly improved amino acid profile (tryptophan increased by 41%, lysine increased by 30%, histidine increased by 23%, and methionine increased by 3.4%), and was released in India in 2008 (Gupta et al. 2009). Liu et al. (2016) identified *q27* as a functional marker associated with endosperm modifications, which has opened up a new avenue for QPM breeding. Shetti et al. (2020) introgressed *o2* allele into parental lines of popular maize hybrids

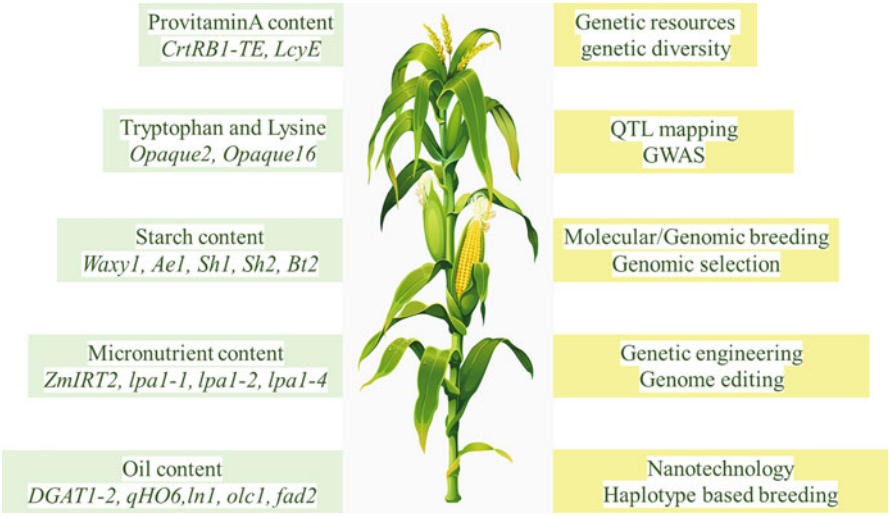


Fig. 1 Genomic designing for nutritionally richmaize

and observed 20–40% higher tryptophan content in converted lines. Sarika et al. (2018) investigated F_2 populations derived from crossing normal maize lines (CML533 and CML537) with a mutant *o16* donor line (QCL3024) and observed that the *o16o16* mutant combination results in genotypes with approximately two times the lysine and tryptophan content than of normal maize.

Indian Agriculture Research institute (IARI) have developed several QPM hybrids such as "Pusa HM-4 Improved," "Pusa HM-8 Improved," and "Pusa HM-9 Improved," and released for general cultivation in India in 2017 (Hossain et al. 2018). Four commercial QPM hybrids, HQPM-1, HQPM-4, HQPM-5, and HQPM-7 developed in India via *o16* MABB introgression in their parental lines have 49–60% increase in lysine and tryptophan contents (Sarika et al. 2018). Through MABB, CIMMYT recently developed several QPM versions (CML244Q, CMl246Q, CML349Q, and CML354Q) of popular inbred lines (CML244Q, CMl246Q, CML349Q, and CML354Q). The performance of these QPM versions is comparable to that of their normal versions (Qureshi et al. 2019).

8.2 MAB for Oil Content

The *DGAT1-2* gene has been majorly used in MAS for oil improvement of maize kernels. Hao et al. (2014) used MABC to transfer the *DGAT1-2* from the high-oil inbred line (By804) into two parents of Zhengdan958 and effectively improved the oil content without changing grain weight. The major QTL *qHO6* for enhancing oil content in maize grain located on chromosome 6 has also been extensively used in MABB (Hao et al. 2014) and developed high-oil populations and hybrids, namely Illinois 6021, 6052, 6001, and Burr white.

8.3 MAB for Starch Content

The discovery of the *ae* (amylose-extender) mutant alleles led to development of high-amylose plants. Amylose content is significantly higher in maize endosperm with the *ae* recessive alleles. Chen et al. (2010) described the use of *ae* (amylose-extender) recessive mutant alleles to create high-amylose cultivars and able to detect *ae* alleles in a backcross and its second generation with this marker more efficiently (53.3 and 73.3%, respectively) than without marker selection.

8.4 MAB for Provitamin A Content

Naturally existing alleles of two key genes of carotenoid pathway were identified, Carotene hydroxylase1 (*CrtRB1*), and lycopene epsilon cyclase *(LcyE*) (Harjes et al. 2008). The favorable allele at *CrtRB1* reduces hydroxylation of β-carotene into β-cryptoxanthin, whereas *LcyE* reduces flux into the α-branch of the pathway. Babu et al. (2013) validated the effects of 3 functional polymorphisms (*LcyE5′TE, LcyE3′Indel* and *CrtRB1-3′TE*) and observed around twofold to tenfold increase in β-carotene (BC) and total provitamin A (proA) content. The *CrtRB1-3′TE* allele have large, significant effect on enhancing BC and total ProA content, irrespective of genetic constitution for *LcyE5′TE* and genetic background (Babu et al. 2013; Sagare et al. 2019), and the favourable*CrtRB1* allele is more effective in increasing PVA content than the favorable *LcyE* allele (Babu et al. 2013). Muthusamy et al. (2014) introgressed favorable allele of *CrtRB-1* into seven elite inbreds and observed β-carotene concentration varied from 8.6 to 17.5 μg/g among the *crtRB1*-introgressed inbreds which was 12.6-fold higher than original lines. Several inbred lines and hybrids have been developed in recent past possessing favorable alleles of *CrtRB-1* and *LcyE* through molecular breeding to enhance provitamin A content (Babu et al. 2013; Azmach et al. 2013; Muthusamy et al. 2015; Yang et al. 2018; Maqbool et al. 2018; Sagare et al. 2019; Mehta et al. 2020; Natesan et al. 2020), and the increase in provitamin A content varied from 5 to 18 μg/g in introgressed lines and hybrids. Recently, Natesan et al. (2020) conducted MABC to transfer the β-carotene gene, *crtRB1*, into UMI1200 and UMI1230 using HP467-15 as the donor parent using one gene-specific marker (*crtRB1 3′TE*), and using the improved lines, five hybrid combinations were developed and ACM-M13-002 was identified as a superior hybrid with a 7.3-fold increase in β-carotene concentration.

8.5 Pyramiding Grain Quality Genes

Gene pyramiding is the process of combining desirable traits by stacking multiple genes into a single genotype. Via molecular breeding approaches several grain quality genes/ QTLs have been pyramided in maize (Table 3) without compromising grain yield. In India, QPM was introgressed with provitamin A, resulting in the provitamin A-enriched elite QPM inbreds CML161 and CML171, Pusa HQPM-5

Table 3 Pyramiding nutritional quality traits in maize through molecular breeding

Traits	Gene combination	Nutritional trait value	Reference
QPM and ProvitaminA	*CrtRB1* in QPM inbreds	5.25–8.14 μg/g provitaminA, 0.35% Lysine	Liu et al. 2015
QPM, ProvitaminA, Vitamin E	*CrtRB1*, *LcyE*, and *VTE4*	16.8% μg/g α-tocopherol, 11.5 μg/g provitaminA, 0.367% Lysine, 0.085% Tryptophan	Hossain et al. 2018
QPM, ProvitaminA, Low phytate	*lpa1-1*, *lpa2-1* in provitamin enriched QPM inbreds	8.3–11.5 μg/g provitaminA, 0.081–0.087% Tryptophan, 0.323–0.372% Lysine, 30–40% reduction in phytic acid phosphorus	Bhatt et al. 2018
QPM and ProvitaminA	*CrtRB1* in QPM inbreds	10.75 μg/g provitaminA, 0.080% Tryptophan, 0.303% Lysine	Goswami et al. 2019
QPM and ProvitaminA	*CrtRB1* in QPM inbreds	6.25–6.80 μg/g provitaminA, 0.080% Tryptophan, 0.334% Lysine	Sagare et al. 2019
QPM and ProvitaminA	opaque2 in β-carotene riched lines	6.12–7.38 μg/g β-carotene, 0.073–0.081% Tryptophan, and 0.294–0.332% Lysine	Chandran et al. 2019
QPM and ProvitaminA	*sh2, opaque2, lcyE and crtRB1* in parental lines of sweetcorn hybrid	18.98 μg/g Provitamin A, 0.39% Tryptophan, 0.10% Lysine, 17.04% brix	Mehta et al. 2020
QPM, Provitamin A, kernel sweetness	*shrunken2, opaque2, lcyE* and *crtRB1* in sweetcorn hybrid	0.390% Lysine, 0.082% tryptophan, 21.14 ppm Provitamin A, 18.96% brix	Baveja et al. 2021
Low amylopectin	waxy1 in white and yellow kernel maize inbreds	96.7% Amylopectin	Talukder et al. 2022
QPM and low amylopectin	*waxy1* and *opaque2* in inbreds	98.84% Amylopectin, 0.102% Tryptophan, and 0.384% Lysine	Talukder et al. 2022

improved, PVCBML6, PVCBML7, Pusa HQPM-7 improved, HKI1128Q, and Pusa Vivek QPM9 improved (Goswami et al. 2019; Liu et al. 2015; Muthusamy et al. 2014; Sagare et al. 2019). Hossain et al. (2018) reported the introgression of QPM with provitamin A and vitamin E, which resulted in the development of improved QPM and provitamin A-rich hybrids (HQPM-1-PV, HQPM-4-PV, HQPM-5-PV, and HQPM-7-PV) with higher α-tocopherol contents. Bhatt et al. (2018) attempted introgression of QPM, provitamin A content, and low phytate content which resulted in the development of improved versions of elite inbreds (HKI161-PV, HKI163-PV, HKI193-1- PV, and HKI193-2- PV) with higher protein quality, higher provitamin A content, and lower phytate content. Chandran et al. (2019) attempted o2 gene introgression (from HKI163 donor) to β-carotene-rich inbred lines (UMI1200β) and improved lines showed higher lysine (0.29–0.33%), tryptophan (0.07–0.08%), and β-carotene (6.12–7.38 μg/g) contents, along with improved agronomic performance. Parental lines of shrunken2 (*sh2*)-based sweet corn hybrids (ASKH-1 and ASKH-2) were subjected to *crtRB1* and *o2* gene introgression through use of

MABB, and reconstituted hybrids with converted lines was found with higher levels of provitamin A (18.98 μg/g), lysine (0.39%), and tryptophan (0.10%), and on par yield as original hybrid (Mehta et al. 2020). Talukder et al. (2022) deployed GAB to pool recessive *waxy1 (wx1)* and *o2* genes in the parental lines of four popular hybrids (HQPM1, HQPM4, HQPM5, and HQPM7) using the gene-based markers. The re-formed hybrids showed 1.4-fold increase in amylopectin and 14.3% and 14.6% increase in lysine and tryptophan respectively over the original hybrids (lysine- 0.336%, tryptophan- 0.089%). Three promising maize hybrids derived from MABB (improved Pusa Vivek Hybrid-27 – provitamin A), Pusa HQPM5 Improved – QPM and provitamin A, and Pusa HQPM7 Improved – QPM and provitamin A) have all been developed and approved for commercial cultivation in India in 2019 (Prasanna et al. 2020). MAS was used to introduce QPM with various nutritional traits such as provitamin A, vitamin E, and low phytate content.

9 Cloning of Grain Quality-Related Genes

The approach of determining the genetic basis of a mutant trait by exploring the linkage with markers whose physical location in the genome is known is referred as map-based cloning, also known as positional cloning. Because of the large amounts of repetitive DNA in maize, positional cloning was thought to be nearly impossible in the last century. However, now that maize physical maps and large number of available markers, and, most importantly, synteny conservation across cereal genomes, make it possible to consider a chromosome walk in less time than cloning by transposon tagging (Bortiri et al. 2006). The following factors pose significant challenges to gene discovery in maize: (1) large genome size, (2) variation in gene order and genome size, (3) a higher rate of multicopy genes, and (4) repetitive sequences and transposons. In maize, the genetic basis of various traits varies greatly. Positional cloning was done for different traits in maize such as disease-resistant, male sterility, fertility restoration, plant architecture etc. But very less work of positional cloning is done on health-related/grain quality traits. Till date several genes related to maize nutritional quality have been cloned such as, opaque1 (*o1*), foury4 (*f4*) and Mucronate (*Mc*) for protein quality, linoleic acid1 (*ln1*), Oleic acid content1 (*olc1*), fatty acid desaturation 2 (*fad2*), and fad6 for oil content, and some of the starch content related genes – Shrunken1 (*Sh1*), *Sh2* and Brittle2 (*Bt2*).

Buckner et al. (1990) cloned the *y1* locus of maize involved in the carotenoid biosynthesis. A maize cDNA encoding phytoene desaturase, an enzyme of the carotenoid biosynthetic pathway was characterized and cloned by Li et al. (1996). The key genes involved in the starch biosynthesis pathway have been cloned, including starch branching enzyme amylose-extender1 (*Ae1*), brittle2 (*Bt2*), Suc synthase shrunken1 (*Sh1*), large subunit of AGPase shrunken2 (*Sh2*), isoamylase-type DBE sugary1 (*Su1*), and granule-bound starch synthase waxy1 (*Wx1*). Zhang et al. (2019) found that wild-type *Se1* is a gene *Zm00001d007657* on chromosome 2 and deletion of this gene causes the sugary enhancer1 (*se1*) phenotype, which is actively involved in starch metabolism in maize endosperm. Recently, Li et al. (2022)

isolated a novel gene from maize (*ZmIRT2*), which exhibited highly homologous to *ZmIRT1*. *ZmIRT2* was expressed in roots and anther and was induced by Fe and zinc (Zn) deficiencies. *ZmIRT2* overexpression in maize led to elevated Zn and Fe levels in roots, shoots, and seeds of transgenic plants. *ZmIRT1* transcript levels were higher in the roots of *ZmIRT2* transgenic plants. According to this study, *ZmIRT2* may only work with *ZmIRT1* to mediate Fe uptake in roots, and *ZmIRT2* could also be used in fortification efforts to increase Zn and Fe levels in crop plants.

10 Recent Concepts and Strategies Developed

10.1 Gene Editing

Genome engineering technologies such as ZFNs (zinc finger nucleases), TALENs (transcription activator-like effector nucleases) and CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats (CRISPR)-associated 9 (Cas9) endonucleases) along with tissue specific expression of functional genes have accelerated molecular/genomic breeding (Chilcoat et al. 2017) Maize has large number of commercialized transgenic events (Yadava et al. 2016), this demonstrates that maize improvement has always been taken prominently and genome engineering is powerful technology to pave the way for enhancing and stacking nutritional quality traits in maize.

The phytate content in maize was reduced by blocking IPK1 gene encoding inositol-1,3,4,5,6-pentakisphosphate 2-kinase using ZFNs (Shukla et al. 2009). Three genes of phytate biosynthetic pathway, *ZmIPK1A*, *ZmIPK*, and *ZmMRP4* were knocked out using CRISPR/Cas9 and TALENS (Liang et al. 2014). RNA interference (RNAi) and CRISPR/Cas9 knockout *ZmMADS47* gene encoding a MADS-box protein, that interact with *o2* to activate zein gene promoter, and the reduced zein content 16.8 and 12.5% were observed in the kernels of *ZmMADS47 RNAi* and MADS/CAS9-21 lines, respectively (Qi et al. 2016). Starch metabolism in maize was altered using CRIPSR/Cas9 to disrupt the waxy gene (*Wx1*), which encodes GBSS being responsible for synthesizing amylose in maize (Waltz 2016). Several nutritional quality traits such as, QPM (*o2, o16*), provitaminA (*CrtRB1, LcyE*), kernel sweetness (*Shrunken2*), low amylopectin (*waxy1*), etc have been well characterized and gene editing could serve as a powerful system for stacking these traits in agronomically superior and high yielding genetic backgrounds (Table 4).

10.2 Nanotechnology

Nanotechnology is a study of nanoparticles (NPs) with unique physical, chemical and biological properties, and in recent past use of nanotechnology in agriculture has been accelerated due to its wide range of applications in crop improvement. The major applications of nanotechnology in agriculture includes; improved seed

Table 4 Potential targets for gene editing to improve nutritional quality in maize

Trait	Genes	Protein function	Gene editing strategy
Starch content	SBE1, SBE3, SBE4, AE1	Starch branching enzymes	Knockout
	SS1, SS2, SS3, SS4, SS5, SS6, SS7, Du1, Su2	Starch synthase	Overexpression
	Waxy1, GBSS1	Granule bound starch synthase	Overexpression
	GPM177	Protein targeting to starch	Overexpression
Oil content	PEP1	Phosphoenolpyruvate carboxylase	Knockout
	ACC1, ACC2, TIDP3607	Acetyl-CoA carboxylase	Overexpression
	LN1, DGATI2	Diacylglycerol acyltransferase	Overexpression
	OLE1, OLE3, OLE4	Delta-9 desaturase	Overexpression
	FAD2	Delta-12 fatty acid desaturase	Knockout
	KCS1, KCS16	Fatty acid elongase	Knockout
Essential amino acid content	VSP1, VSP2, BIP1, BIP2, BIP3	Storage protein	Overexpression
	HNT1	Homocysteine S-methyltransferase	Overexpression
	GOT1, GOT2, GOT3, GOT4	Aspartate aminotransferase	Overexpression
	ASN3, ASN4	Asparagine synthetase	Overexpression
Vitamin content	VP5	Phytoene desaturase	Overexpression
	DXS1	1-deoxyxylulose 5-phosphate synthase	Overexpression
	DXR1, DXR2	Deoxy-D-xylulose 5-phosphate reductoisomerase	Overexpression
	NCED6, NCED8	Carotenoid cleavage dioxygenase	Knockout
	HPT1	Homogentisatephytyltransferase	Overexpression
	SXD1	Tocopherol cyclase	Overexpression
	DHFS1, DHFS2	Dihydrofolate synthetase	Overexpression
	FGP2, BM4	Folylpolyglutamate synthase	Overexpression
Mineral content	NAS1, NAS2, NAS3, NAS4, NAS6, NAS8, NAS9, NAS10	Nicotianamine synthase	Overexpression
	NAAT1, PCO115235C	Nicotianamine aminotransferase	Overexpression
	IDP871	Small GTPase	Overexpression
	MIPS2	Myo-inositol-1-phosphate synthase	Knockout
	PAP2, PAP22	Phytase	Overexpression

germination using nanoformulations, nanofertilizers for nutrient use efficiency in crop that helps in accumulation of nutritional quality components in the edible portion, weed control using herbicides, disease and pest control using nanopesticides and nanosensors, nanopackaging for improving shelf life of produce, nanosensors for post-harvest quality analysis. Research on NPs for improving nutrition quality-

related traits in cereals has gained attention in the recent past (Bajpai et al. 2020). In the agronomic biofortification approach, application of nanofertilizers to enhance nutrient use efficiency of crop and subsequently to improve nutritional quality of maize is an emerging approach. Surface-modified zeolite enhances phosphate and sulfate uptake (Li and Zhang 2010), Montmorillonite and Zeolite improves nitrogen uptake via xylem (Manikandan and Subramanian 2014), Nanocomposites containing organic polymer intercalated in the layers of kaolinite clays increases uptake of wide range of micronutrients. Recent developments in genetic engineering and engineered nanomaterials based targeted delivery of CRISPR/Cas mRNA, and sgRNA for the genetic modification of crops is a noteworthy scientific achievement and this will pave the way to enhance the nutritional content of major cereals like rice, corn, and wheat.

11 Genetic Engineering for Nutritional Quality Traits

Biosynthetic and metabolic pathways of nutritional traits of maize are well studied. Genetic engineering (GE) approach provides opportunity to enhance nutrition content of maize via several approaches viz., transformation of single gene, transformation with multiple genes (encoding different nutrients) in a single cassette, overexpression of nutrient transporters, silencing of feedback inhibition enzymes of nutrient biosynthesis pathways, overexpression of major enzymes of pathways. Genetic engineering for enhancing nutrient content of crops by modifying biosynthesis pathways is called metabolic engineering.

11.1 GE to Enhance Essential Amino Acids

Zein are the major proteins in maize and deficit in essential amino acids lysine and tryptophan. T-DNA mediated mutations in 19- and 22-kD α-zeins in maize enhanced seed Lys and Trp content and reduced zein levels. A genetically engineered high lysine maize (0.360% vs. 0.255% in common maize kernel), LY038 was developed by transformation of the *cordapA* gene from *Corynobacterium glutamicum*, a soil bacterium, into the maize genome (Lucas et al. 2007). LY038 is the first genetically modified (GM) maize, approved for commercial use in countries like Colombia, Mexico, Canada, the Philippines, the USA, and Japan (EPA 2020). Tryptophan synthesis is feedback regulated by inhibiting Anthranilate synthase (AS), and 1.1- to 50-fold increase in free amino acid content in maize transformed with feedback altered AS gene than in non-transformed maize is reported.

11.2 GE to Enhance Micronutrients

Phytic acid (phytate) is a chelator, reduces bioavailability of essential micronutrients such as Fe, Zn, Ca, Cu, Mn etc. The *low phytic acid* (*lpa*) mutant of

maize viz., *lpa1-1* and *lpa2-1*, *lpa241* were found with 50–90% reduction in phytate, but have negative effect on seed germination and plant growth. Genetic modification of *lpa1-1* maize by endosperm specific overexpression of soybean ferritin gene was resulted in more than twofold improvement in iron bioavailability, and transgenic *lpa1-1* seeds were having higher germination rates and seedling vigor compared to non-transgenic seeds (Aluru et al. 2011). A significant enhancement in bioavailable Fe was observed in transgenic maize lines generated by endosperm specific co-expression of recombinant Soybean ferritin and *Aspergillus* Phytase genes.

11.3 GE to Enhance Carotenoids

Multiplex transgenic maize developed by transferring 5 carotenogenic genes, *Zmpsy1* (*Zea mays phytoene synthase 1*), *PacrtI* (*Pantoeaananatis phytoene desaturase*), *Gllycb* (*Gentiana lutea lycopene-cyclase*), *Glbch* (*Gentiana lutea carotene hydroxylase*), and *ParacrtW* (*Paracoccus*carotene ketolase) driven by different endosperm-specific promoters have resulted in 50- to 60-fold increase in total carotenoids. Carotenoid intermediates such as, keto-carotenoids, asthaxanthi, adonixanthin, 3-hydroxye chinenone, echinenone were engineered in transgenic maize. Genetic modification in maize by transferring *ZmPSY1 cDNA* (with LMW glutenin promoter from wheat), *Crt1* gene from *Pantoeaananatis* (with barley Dhordein promoter), *Osdhar* gene (for ascorbate) and E. coli *folE* gene (for folate) were found to possess 169-fold higher β-carotene, 6-fold higher ascorbate, and double amount of folate than non-transformed maize (Naqvi et al. 2009). Transgenic maize produced by overexpression of *Psy*gene and silencing *LcyE* gene resulted in diverting carotenoid biosynthesis pathway toward β-branch, and thus, with increasing high value astaxanthin carotenoid in the kernel (Farre et al. 2016). Overexpression of the *IbOr* gene in maize inbred lines under the control of maize seed-specific promoter globulin 1 (*Glo1*) resulted in 10.36- and 15.11-folds increase in total carotenoid and ß-carotene contents (Tran et al. 2017).

11.4 GE to Increase Oil and Starch Content

Maize transformed with *Puroindoline* genes (*Pina* and *pinb*) from wheat resulted in 25.23% increase in a total oil content (Zhang et al. 2010). Overexpression of *ZmWRI1* (*WRINKLED1*) gene in maize resulted in enhancing oil content without affecting seed germination, plant growth and grain yield (Shen et al. 2010). The maize transformed with fungal *diacylglycerol acyltransferase2* (*DGAT2*) genes from *Umbelopsisramanniana* and *Neurospora crassa* under the control of embryo enhanced promoter, showed 26% increase in kernel oil content compared to non-transgenic maize (Oakes et al. 2011).

11.5 GE to Increase Starch Content

Genetic transformation of maize with *Bt2* and *Sh2* genes from wild maize under the control of endosperm specific promoter resulted in increased starch content in transformed lines (Li et al. 2008). Overexpression of *Bt2*, *Sh2*, *Sh1* and *GbssIIa* genes, and silencing *SbeI* and *SbeIIb* by RNA interference to enhance activity of sucrose synthase, AGPase and granule-bound starch synthase, and to reduce the activity of starch branching enzyme in maize, resulted in 2.8–7.7% increase in kernel starch and 37.8–43.7% increase in amylose content (Jiang et al. 2013). The overexpression of the mutated *ZmDA1* and *ZmDAR1* genes driven by maize ubiquitin promoter resulted in improving the sugar imports into the sink organ and starch synthesis in maize kernels (Xie et al. 2018).

12 Role of Bioinformatics in Maize Metabolome Improvement

Maizeis a model cereal crop in omics research programs across the world, and studies on maize using multi-omics datasets could help to understand more about domestication, selection, and evolution of maize over the period of time. In 2009 the first maize reference sequence B73 was released, further using high-density SNPs several maize lines and population were re-sequenced and mapped against B73 reference genome to expandnutritional related genomics such as oil and vitamin metabolism (Wang et al. 2018). Currently in maize genome database several additional maize genome sequences have been reported such as, PH207, Mexicana, Mo17, W22, HZS, and SK. For successful maize breeding and improvement programs timely utilization of multi-omics tools, datasets, and web-resources could be most imperative (Table 5) These genome sequences, datasets and omics platforms are playing crucial role in maize metabolome improvement, and assisting in diversified breeding purposes to bred varieties with high yield and nutritional quality

13 Conclusion and Future Perspectives

The early generation crop breeding approaches such as, phenotypic selection, cross breeding, mutation breeding have been developed several high yielding, biotic and abiotic stress-resistant varieties and hybrids. And, also identified diversity and resources for traits contributing nutritional quality. These conventional/ traditional breeding approaches are tedious and time consuming and creates difficulty in breeding quantitative traits. The recent advances in molecular tools, techniques, genome sequencing technologies, novel genomic tools such as trait mapping, identifying trait specific major QTLs and genes, haplotype-based allele mining, enriched bioinformatic databases, rapid generation advancement through speed breeding and

Table 5 List of important maize databases and web-resource to obtain genomic information on nutritional quality

Database/ Web-resource	Database/ Web-resource and Establish/Launch	Website
NCBI	National Center for Biotechnology Information	https://www.ncbi.nlm.nih.gov/
MaizeGDB	Maize Genetics and Genomics Database – Lawrence and colleagues, 2003	https://www.maizegdb.org/
SNPversity	MaizeGDB team and	https://www.maizegdb.org/snpversity
MaizeMine	MaizeGDB team and the Elsik lab at the University of Missouri	https://maizemine.rnet.missouri.edu/maizemine/begin.do
MaizeDIG	The Maize Database of Images and Genomes Carson M Andorf and team in 2019	https://maizedig.maizegdb.org/
qTeller	Professor James Schnable from University of Nebraska-Lincoln	https://qteller.maizegdb.org/
PedigreeNet	MaizeGDB- PedigreeNet MaizeGDB team	https://www.maizegdb.org/breeders_toolbox
ZmGDB	*Zea mays* Genome DB – PlantGDB Developed as a part of NSF-funded project "Cyberinfrastructure for (Comparative) Plant Genome Research Through PlantGDB" 2006	https://www.plantgdb.org/ZmGDB/
Panzea	NSF-PGRP	http://www.panzea.org
MEILAM	Maize Elite Inbred Line Ac/Ds Mutant database	http://www.maizetepolymorphism.com/AcDs/
Maize FLcDNA	Full Length cDNA Project	http://www.maizecdna.org/
CSRDB	Cereal small RNAs Database	http://sundarlab.ucdavis.edu/smrnas/
ZEAMAP	Maize and its wild relative's database	http://www.zeamap.com/
MaizeSNPDB	MaizeSNP Database – Wen Yao, from College of Life Sciences, Henan Agricultural University	https://venyao.xyz/MaizeSNPDB/
PPIM	Protein-Protein Interaction Database for maize	http://comp-sysbio.org/ppim/
Gramene	Establishment and Maintenance – "Oregon State University, Cold Spring Harbor Laboratory and EMBL-EBI"	https://www.gramene.org/
cropPAL	The compendium of crop Proteins with Annotated Locations Initiated – Cornelia Hooper and Colleagues from The University of Western Australia, Crawley, Australia	https://crop-pal.org/

double haploid, genomic selection, genome editing, nanotechnology, machine learning, and artificial intelligence has led to current generation crop breeding such as, biotech breeding and design breeding. The pathways of nutritional traits are well studies, the information on key enzymes, associated genes/QTLs, genetic diversity

and resources is publicly available. The genome editing technologies and speed breeding approaches are assisting haplotype-based breeding/design breeding which can assist in developing tailor-made maize with higher nutrition.

References

Aluru MR, Rodermel SR, Reddy MB (2011) Genetic modification of low phytic acid 1-1 maize to enhance iron content and bioavailability. J Agric Food Chem 59(24):12954–12962. https://doi.org/10.1021/jf203485a

Amusan IO, Rich PJ, Menkir A, Housley T, Ejeta G (2008) Resistance to Striga hermonthica in a maize inbred line derived from Zea diploperennis. New Phytol 178(1):157–166. https://doi.org/10.1111/j.1469-8137.2007.02355.x

Azmach G, Gedil M, Menkir A, Spillane C (2013) Marker-trait association analysis of functional gene markers for provitamin a levels across diverse tropical yellow maize inbred lines. BMC Plant Biol 13:227. https://doi.org/10.1186/1471-2229-13-227

Azmach G, Menkir A, Spillane C, Gedil M (2018) Genetic loci controlling carotenoid biosynthesis in diverse tropical maize lines. G3 (Bethesda) 8(3):1049–1065. https://doi.org/10.1534/g3.117.300511

Babu R, Prasanna BM (2014) Molecular breeding for quality protein maize (QPM). In: Tuberosa R, Graner A, Frison E (eds) Genomics of plant genetic resources. Springer, Dordrecht, pp 489–505

Babu R, Rojas NP, Gao S, Yan J, Pixley K (2013) Validation of the effects of molecular marker polymorphisms in LcyE and CrtRB1 on provitamin a concentrations for 26 tropical maize populations. Theor Appl Genet 126(2):389–399. https://doi.org/10.1007/s00122-012-1987-3

Babu BK, Agrawal PK, Saha S, Gupta HS (2015) Mapping QTLs for opaque2 modifiers influencing the tryptophan content in quality protein maize using genomic and candidate gene-based SSRs of lysine and tryptophan metabolic pathway. Plant Cell Rep 34(1):37–45. https://doi.org/10.1007/s00299-014-1685-5

Bajpai A, Jadhav K, Muthukumar M, Kumar S, Srivatava G (2020) Use of nanotechnology in quality improvement of economically important agricultural crops. In: Ghorbanpour M, Bhargava P, Varma A, Choudhary DK (eds) Biogenic nano-particles and their use in agro-ecosystems. Springer, Singapore, pp 39–57

Barker NP, Clark LG, Davis JI, Duvall MR, Guala GF, Hsiao C et al (2001) Grass phylogeny working group G. phylogeny and subfamilial classification of the grasses (Poaceae). Ann Mo Bot Gard 88(3):373–457

Baveja A, Muthusamy V, Panda KK, Zunjare RU, Das AK, Chhabra R, Mishra SJ, Mehta BK, Saha S, Hossain F (2021) Development of multinutrient-rich biofortified sweet corn hybrids through genomics-assisted selection of shrunken2, opaque2, lcyE and crtRB1 genes. J Applied Genet 62:419–429

Bhatt V, Muthusamy V, Jha S, Zunjare RU, Baveja A, Dosad S (2018) Development of low phytic acid maize through marker assisted introgression of *lpa1-1* and *lpa2-1* genes. In: Admin S (ed) 13th Asian maize conference on and expert consultation on maize for food, feed, nutrition and environmental security, Ludhiana, India, October 8–10, pp 143–144

Bortiri E, Chuck G, Vollbrecht E, Rocheford T, Martienssen R, Hake S (2006) ramosa2 encodes a LATERAL ORGAN BOUNDARY domain protein that determines the fate of stem cells in branch meristems of maize. Plant Cell 18(3):574–585. https://doi.org/10.1105/tpc.105.039032

Buckner B, Kelson TL, Robertson DS (1990) Cloning of the y1 locus of maize, a gene involved in the biosynthesis of carotenoids. Plant Cell 2(9):867–876. https://doi.org/10.1105/tpc.2.9.867

Campbell MR, Jane JL, Pollak L, Blanco M, O'Brien A (2007) Registration of maize germplasm line GEMS-0067. J Plant Regist 1(1):60–61. https://doi.org/10.3198/jpr2006.10.0640crg

Chandran S, Pukalenthy B, Adhimoolam K, Manickam D, Sampathrajan V, Chocklingam V et al (2019) Marker-assisted selection to pyramid the opaque-2 (o2) and β-carotene (crtRB1) genes in maize. Front Genet 10:859. https://doi.org/10.3389/fgene.2019.00859

Chen F, Zhu SW, Xiang Y, Jiang HY, Cheng BJ (2010) Molecular marker-assisted selection of the ae alleles in maize. Genet Mol Res 9(2):1074–1084. https://doi.org/10.4238/vol9-2gmr799

Chilcoat D, Liu ZB, Sander J (2017) Use of CRISPR/Cas9 for crop improvement in maize and soybean. Prog Mol Biol Transl Sci 149:27–46. https://doi.org/10.1016/bs.pmbts.2017.04.005

Cook JP, McMullen MD, Holland JB, Tian F, Bradbury P, Ross-Ibarra J et al (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol 158(2):824–834.

Das AK, Jaiswal SK, Muthusamy V, Zunjare RU, Chauhan HS, Chand G et al (2019) Molecular diversity and genetic variability of kernel tocopherols among maize inbreds possessing favourable haplotypes of γ-tocopherol methyl transferase (ZmVTE4). J Plant Biochem Biotechnol 28(3):253–262. https://doi.org/10.1007/s13562-018-0470-x

Deng M, Li D, Luo J, Xiao Y, Liu H, Pan Q et al (2017) The genetic architecture of amino acids dissection by association and linkage analysis in maize. Plant Biotechnol J 15(10):1250–1263. https://doi.org/10.1111/pbi.12712

Environmental Protection Agency (EPA) (2020) Predicting the impact of coexistence-guided, genetically modified cropping on Irish biodiversity. STRIVE Environmental Protection Agency Program, Ireland, pp 2007–2013

Farré G, Perez-Fons L, Decourcelle M, Breitenbach J, Hem S, Zhu C et al (2016) Metabolic engineering of astaxanthin biosynthesis in maize endosperm and characterization of a prototype high oil hybrid. Transgenic Res 25(4):477–489. https://doi.org/10.1007/s11248-016-9943-7

Goswami R, Zunjare RU, Khan S, Baveja A, Muthusamy V, Hossain F (2019) Marker-assisted introgression of rare allele of β-carotene hydroxylase (crtRB1) gene into elite quality protein maize inbred for combining high lysine, tryptophan and provitamin A in maize. Plant Breed 138 (2):174–183. https://doi.org/10.1111/pbr.12676

Gupta HS, Agrawal PK, Mahajan V, Bisht GS, Kumar A, Verma P et al (2009) Quality protein maize for nutritional security: rapid development of short duration hybrids through molecular marker assisted breeding. Curr Sci 96(2):230–237

Hao X, Li X, Yang X, Li J (2014) Transferring a major QTL for oil content using marker-assisted backcrossing into an elite hybrid to increase the oil content in maize. Mol Breeding 34(2):739–748. https://doi.org/10.1007/s11032-014-0071-x

Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG et al (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science 319 (5861):330–333. https://doi.org/10.1126/science.1150255

Hindu V, Palacios-Rojas N, Babu R, Suwarno WB, Rashid Z, Usha R et al (2018) Identification and validation of genomic regions influencing kernel zinc and iron in maize. Theor Appl Genet 131 (7):1443–1457. https://doi.org/10.1007/s00122-018-3089-3

Holding DR (2014) Recent advances in the study of prolamin storage protein organization and function. Front Plant Sci 5:276. https://doi.org/10.3389/fpls.2014.00276

Hossain F, Muthusamy V, Pandey N, Vishwakarma AK, Baveja A, Zunjare RU et al (2018) Marker-assisted introgression of opaque2 allele for rapid conversion of elite hybrids into quality protein maize. J Genet 97(1):287–298. https://doi.org/10.1007/s12041-018-0914-z

Ivanov K, Tonev T, Nguyen N, Peltekov A, Mitkov A (2019) Impact of foliar fertilization with nanosized zinc hydroxy nitrate on maize yield and quality. Emir J Food Agric 31:597–604. https://doi.org/10.9755/ejfa.2019.v31.i8.2003

Jaiswal SK, Muthusamy V, Hossain F, Zunjare RU, Goswami R, Chhabra R et al (2019) Characterization of maize genotypes using microsatellite markers associated with QTLs for kernel iron and zinc. Indian J Biotechnol 18:224–234

Jaradat AA, Goldstein W (2013) Diversity of maize kernels from a breeding program for protein quality: I. physical, biochemical, nutrient, and color traits. Crop Sci 53(3):956–976. https://doi.org/10.2135/cropsci2012.07.0437

Jiang L, Yu X, Qi X, Yu Q, Deng S, Bai B et al (2013) Multigene engineering of starch biosynthesis in maize endosperm increases the total starch content and the proportion of amylose. Transgenic Res 22(6):1133–1142. https://doi.org/10.1007/s11248-013-9717-4

Karn A, Gillman JD, Flint-Garcia SA (2017) Genetic analysis of teosinte alleles for kernel composition traits in maize. G3 (Bethesda) 7(4):1157–1164. https://doi.org/10.1534/g3.117.039529

Khan NU, Sheteiwy M, Lihua N, Khan MMU, Han Z (2019) An update on the maize zein-gene family in the post-genomics era. Food Prod Process Nutr 1(1):1–12

Kumar A, Kumari J, Rana JC, Chaudhary DP, Kumar R, Singh H et al (2015) Diversity among maize landraces in north west Himalayan region of India assessed by agro-morphological and quality traits. Ind J Gen Plant Breed 75(2):188–195. https://doi.org/10.5958/0975-6906.2015.00029.2

Lambert R (2001) High-oil corn hybrids. In: Hallauer A (ed) Specialty corns, 2nd edn. CRC Press, Boca Raton, pp 131–154

Lambert RJ, Alexander DE, Han ZJ (1998) A high oil pollinator enhancement of kernel oil and effects on grain yields of maize hybrids. Agron J 90(2):211–215. https://doi.org/10.2134/agronj1998.00021962009000020016x

Lennon JR, Krakowsky M, Goodman M, Flint-Garcia S, Balint-Kurti PJ (2016) Identification of alleles conferring resistance to gray leaf spot in maize derived from its wild progenitor species teosinte. Crop Sci 56(1):209–218. https://doi.org/10.2135/cropsci2014.07.0468

Li Z, Zhang Y (2010) Use of surfactant-modified zeolite to carry and slowly release sulfate. Desalin Water Treat 21(1–3):73–78. https://doi.org/10.5004/dwt.2010.1226

Li ZH, Matthews PD, Burr B, Wurtzel ET (1996) Cloning and characterization of a maize cDNA encoding phytoene desaturase, an enzyme of the carotenoid biosynthetic pathway. Plant Mol Biol 30(2):269–279. https://doi.org/10.1007/BF00020113

Li L, Jiang H, Campbell M, Blanco M, Jane JL (2008) Characterization of maize amylose-extender (ae) mutant starches. Part I: relationship between resistant starch contents and molecular structures. Carbohydr Polym 74(3):396–404. https://doi.org/10.1016/j.carbpol.2008.03.012

Li H, Peng Z, Yang X, Wang W, Fu J, Wang J et al (2013) Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat Genet 45(1):43–50

Li C, Huang Y, Huang R, Wu Y, Wang W (2018) The genetic architecture of amylose biosynthesis in maize kernel. Plant Biotechnol J 16(2):688–695. https://doi.org/10.1111/pbi.12821

Li S, Song Z, Liu X, Zhou X, Yang W, Chen J et al (2022) Mediation of zinc and iron accumulation in maize by ZmIRT2, a novel iron-regulated transporter. Plant Cell Physiol 63(4):521–534. https://doi.org/10.1093/pcp/pcab177

Liang Z, Zhang K, Chen K, Gao C (2014) Targeted mutagenesis in Zea mays using TALENs and the CRISPR/Cas system. J Genet Genomics 41(2):63–68. https://doi.org/10.1016/j.jgg.2013.12.001

Lin F, Zhou L, He B, Zhang X, Dai H, Qian Y et al (2019) QTL mapping for maize starch content and candidate gene prediction combined with co-expression network analysis. Theor Appl Genet 132(7):1931–1941. https://doi.org/10.1007/s00122-019-03326-z

Liu L, Jeffers D, Zhang Y, Ding M, Chen W, Kang MS et al (2015) Introgression of the crtRB1 gene into quality protein maize inbred lines using molecular markers. Mol Breed 35(8):154. https://doi.org/10.1007/s11032-015-0349-7

Liu H, Shi J, Sun C, Gong H, Fan X, Qiu F et al (2016) Gene duplication confers enhanced expression of 27-kDa γ-zein for endosperm modification in quality protein maize. Proc Natl Acad Sci U S A 113(18):4964–4969. https://doi.org/10.1073/pnas.1601352113

Lucas DM, Taylor ML, Hartnell GF, Nemeth MA, Glenn KC, Davis SW (2007) Broiler performance and carcass characteristics when fed diets containing lysine maize (LY038 or LY038× MON 810), control, or conventional reference maize. Poult Sci 86(10):2152–2161. https://doi.org/10.1093/ps/86.10.2152

Manikandan A, Subramanian KS (2014) Fabrication and characterisation of nanoporous zeolite based N fertilizer. Afr J Agric Res 9(2):276–284.

Maqbool MA, Aslam M, Beshir A, Khan MS (2018) Breeding for provitamin A biofortification of maize (Zea mays L.). Plant Breed 137(4):451–469. https://doi.org/10.1111/pbr.12618

Mehta BK, Muthusamy V, Zunjare RU, Baveja A, Chauhan HS, Chhabra R et al (2020) Biofortification of sweet corn hybrids for provitamin-A, lysine and tryptophan using molecular breeding. J Cereal Sci 96:103093. https://doi.org/10.1016/j.jcs.2020.103093

Mir C, Zerjal T, Combes V, Dumas F, Madur D, Bedoya C et al (2013) Out of America: tracing the genetic footprints of the global diffusion of maize. Theor Appl Genet 126(11):2671–2682. https://doi.org/10.1007/s00122-013-2164-z

Muthusamy V, Hossain F, Thirunavukkarasu N, Choudhary M, Saha S, Bhat JS et al (2014) Development of β-carotene rich maize hybrids through marker-assisted introgression of β-carotene hydroxylase allele. PLOS ONE 9(12):e113583

Muthusamy V, Hossain F, Thirunavukkarasu N, Saha S, Gupta HS (2015) Allelic variations for lycopene-ε-cyclase and β-carotene hydroxylase genes in maize inbreds and their utilization in β-carotene enrichment programme. Cogent Food Agric 1(1):1033141. https://doi.org/10.1080/23311932.2015.1033141

Naqvi S, Zhu C, Farre G, Ramessar K, Bassie L, Breitenbach J et al (2009) Transgenic multivitamin corn through biofortification of endosperm with three vitamins representing three distinct metabolic pathways. Proc Natl Acad Sci U S A 106(19):7762–7767. https://doi.org/10.1073/pnas.0901412106

Natesan S, Duraisamy T, Pukalenthy B, Chandran S, Nallathambi J, Adhimoolam K et al (2020) Enhancing β-carotene concentration in parental lines of CO6 maize hybrid through marker-assisted backcross breeding (MABB). Front Nutr 7:134. https://doi.org/10.3389/fnut.2020.00134

Oakes J, Brackenridge D, Colletti R, Daley M, Hawkins DJ, Xiong H et al (2011) Expression of fungal diacylglycerol acyltransferase2 genes to increase kernel oil in maize. Plant Physiol 155 (3):1146–1157. https://doi.org/10.1104/pp.110.167676

Owens BF, Lipka AE, Magallanes-Lundback M, Tiede T, Diepenbrock CH, Kandianis CB et al (2014) A foundation for provitamin A biofortification of maize: genome-wide association and genomic prediction models of carotenoid levels. Genetics 198(4):1699–1716. https://doi.org/10.1534/genetics.114.169979

Prasanna BM, Vasal SK, Kassahun B, Singh NN (2001) Quality protein maize. Curr Sci 81(10): 1308–1319

Prasanna BM, Palacios-Rojas N, Hossain F, Muthusamy V, Menkir A, Dhliwayo T et al, San Vicente F, Nair SK, Vivek BS, Zhang (2020) Molecular breeding for nutritionally enriched maize: status and prospects. Front Genet X:1392

Qi W, Zhu T, Tian Z, Li C, Zhang W, Song R (2016) High-efficiency CRISPR/Cas9 multiplex gene editing using the glycine tRNA-processing system-based strategy in maize. BMC Biotechnol 16 (1):58. https://doi.org/10.1186/s12896-016-0289-2. PMID 27515683

Qureshi AMI, Dar ZA, Wani SH (eds) (2019) Quality breeding in field crops. Springer International Publishing, Cham

Risliawati A, Trikoesoemaningtyas S, Suwarno WB, Lestari P (2022) Existing diversity profile for kernel characteristics of maize germplasm in IAARD-ICABIOGRAD gene bank. AIP Conf Proc 2462(1):020007

Sagare DB, Shetti P, Surender M, Reddy SS, Pradeep T, Anuradha G (2018) Maize: potential crop for provitamin A biofortification. Maydica 63(2):11

Sagare DB, Shetti P, Surender M, Reddy SS (2019) Marker-assisted backcross breeding for enhancing β-carotene of QPM inbreds. Mol Breed 39(2):1–12. https://doi.org/10.1007/s11032-019-0939-x

Sarika K, Hossain F, Muthusamy V, Zunjare RU, Baveja A, Goswami R et al (2018) Marker-assisted pyramiding of opaque2 and novel opaque16 genes for further enrichment of lysine and tryptophan in sub-tropical maize. Plant Sci 272:142–152. https://doi.org/10.1016/j.plantsci.2018.04.014

Shen B, Allen WB, Zheng P, Li C, Glassman K, Ranch J et al (2010) Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize. Plant Physiol 153(3):980–987. https://doi.org/10.1104/pp.110.157537

Shetti P, Sagare DB, Surender M, Reddy SS (2020) Development of lysine and tryptophan rich maize (Zea mays) inbreds employing marker assisted backcross breeding. Plant Gene 23: 100236. https://doi.org/10.1016/j.plgene.2020.100236

Shukla VK, Doyon Y, Miller JC, DeKelver RC, Moehle EA, Worden SE, Mitchell JC, Arnold NL, Gopalan S, Meng X, Choi VM (2009) Precise genome modification in the crop species Zea mays using zinc-finger nucleases. Nature 459(7245):437–441

Suwarno WB, Pixley KV, Palacios-Rojas N, Kaeppler SM, Babu R (2015) Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. Theor Appl Genet 128 (5):851–864. https://doi.org/10.1007/s00122-015-2475-3

Swarup S, Cargill EJ, Crosby K, Flagel L, Kniskern J, Glenn KC (2021) Genetic diversity is indispensable for plant breeding to improve crops. Crop Sci 61(2):839–852. https://doi.org/10.1002/csc2.20377

Talukder ZA, Muthusamy V, Chhabra R, Gain N, Reddappa SB, Mishra SJ et al (2022) Combining higher accumulation of amylopectin, lysine and tryptophan in maize hybrids through genomics-assisted stacking of waxy1 and opaque2 genes. Sci Rep 12(1):1–16

Teklewold A, Gissa DW, Tadesse A, Tadesse B, Bantte K, Friesen D et al (2015) Quality protein maize (QPM): a guide to the technology and its promotion in Ethiopia, CIMMYT: Addis Ababa, Ethiopia

Tian QZ, Li XH, Li MS, Jiang W, Zhang SH (2004) Molecular markers assisted selection to quality protein maize. J Maize Sci 12(2):108–110

Tran TL, Ho TH, Nguyen DT (2017) Overexpression of the IbOr gene from sweet potato (Ipomea batatas "Hoang Long") in maize increases total carotenoid and β-carotene contents. Turk J Biol 41(6):1003–1010. https://doi.org/10.3906/biy-1708-87

Waltz E (2016) CRISPR-edited crops free to enter market, skip regulation. Nat Biotechnol 34(6): 582. https://doi.org/10.1038/nbt0616-582

Wang L, Xu C, Qu M, Zhang J (2008) Kernel amino acid composition and protein content of introgression lines from *Zea mays* ssp. *mexicana* into cultivated maize. J Cereal Sci 48(2):387–393. https://doi.org/10.1016/j.jcs.2007.09.014

Wang H, Xu S, Fan Y, Liu N, Zhan W, Liu H et al (2018) Beyond pathways: genetic dissection of tocopherol content in maize kernels by combining linkage and association analyses. Plant Biotechnol J 16(8):1464–1475. https://doi.org/10.1111/pbi.12889

Wu F, Zhang S, Gao Q, Liu F, Wang J, Wang X (2021) Genetic diversity and population structure analysis in a large collection of Vicia amoena in China with newly developed SSR markers. BMC Plant Biol 21(1):544. https://doi.org/10.1186/s12870-021-03330-w

Xie G, Li Z, Ran Q, Wang H, Zhang J (2018) Over-expression of mutated Zm DA 1 or Zm DAR 1 gene improves maize kernel yield by enhancing starch synthesis. Plant Biotechnol J 16(1): 234–244. https://doi.org/10.1111/pbi.12763

Yadava P, Abhishek A, Singh R, Singh I, Kaul T, Pattanayak A et al (2016) Advances in maize transformation technologies and development of transgenic maize. Front Plant Sci 7:1949. https://doi.org/10.3389/fpls.2016.01949

Yang R, Yan Z, Wang Q, Li X, Feng F (2018) Marker-assisted backcrossing of lcyE for enhancement of proA in sweet corn. Euphytica 214(8):1–12. https://doi.org/10.1007/s10681-018-2212-5

Zhang J, Martin JM, Beecher B, Lu C, Hannah LC, Wall ML et al (2010) The ectopic expression of the wheat puroindoline genes increase germ size and seed oil content in transgenic corn. Plant Mol Biol 74(4–5):353–365. https://doi.org/10.1007/s11103-010-9679-3

Zhang X, von Mogel KJHV, Lor VS, Hirsch CN, De Vries B, Kaeppler HF et al (2019) Maize sugary enhancer1 (se1) is a gene affecting endosperm starch metabolism. Proc Natl Acad Sci U S A 116(41):20776–20785. https://doi.org/10.1073/pnas.1902747116

Zheng Y, Yuan F, Huang Y, Zhao Y, Jia X, Zhu L et al (2021) Genome-wide association studies of grain quality traits in maize. Sci Rep 11(1):9797

Barley: From Molecular Basis of Quality to Advanced Genomics-Based Breeding

Franca Finocchiaro, Valeria Terzi, and Stefano Delbono

Contents

Abstract

Barley (*Hordeum vulgare* L.) is a cereal crop that belongs to the family of Poaceae (Gramineae) and tribe Triticeae. This cereal ranks fourth in the world production of cereals. Barley is one of the cereals having greater genetic diversity. It is largely used for animal feed and malting, with only a minor part used for human consumption. Barley can be considered an excellent cereal to produce functional foods, thanks above all to the presence of soluble fiber (β-glucans). In barley grains, it is possible to find a lot of bioactive compounds: from carbohydrate polymers (like β-glucan, arabinoxylan, and hemicellulose), to vitamin E (tocopherol and tocotrienols), to secondary metabolites like phenols. Over the

F. Finocchiaro (✉) · V. Terzi · S. Delbono
Council for Agricultural Research and Economics (CREA), Research Centre for Genomics & Bioinformatics, Fiorenzuola d'Arda (PC), Italy
e-mail: franca.finocchiaro@crea.gov.it

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_4

years, breeding programs have therefore been implemented to improve the content of bioactive compounds in barley. Genomics has developed very quickly in the last years, providing tools and technologies that help us to lend precision and efficiency to barley breeding programs for the development of new varieties with healthier properties and resilient to climate changes.

Keywords

Barley · *Hordeum vulgare* · β-glucans · Polyphenols · Tocols · Breeding for HR molecules

1 Introduction

1.1 Agricultural Importance of the Crop

Barley (*Hordeum vulgare* L.) is a cereal crop that belongs to the family of Poaceae (Gramineae) and tribe Triticeae (Fig. 1).

This cereal is the fourth in the world production of cereals, after wheat, maize, and rice. It is a crop that can grow under a wide environmental condition, stress tolerant, and, for these reasons, barley is particularly used in some developing countries that present arid climate condition, being the most important staple food resource. Currently, barley is grown on 48 million hectares in moderate, continental, and subtropical climates. It is mainly employed for animal feeding, to produce beer and

Fig. 1 Barley spike

spirits, and to an ever-increasing extent, for human consumption. The barley world production was about 147.05 million metric tons in the 2021/2022 crop year, with a reduction of about 14.0 million metric tons compared to 2020/2021 (https://www.statista.com/statistics/271973/world-barley-production-since-2008/). Besides being an agronomically relevant crop, barley is considered a very good model for the Triticeae, especially bread, which possess very complex polyploid genomes. Barley is diploid with a genome size of about 5.3 Gb distributed over seven chromosomes.

Barley was one of the first cereals being cultivated: archaeological finds have been found in areas corresponding to the Fertile Crescent and its domestication has been traced back to about 8000 years BC (Giraldo et al. 2019). Since its domestication, barley was a staple food, with considerable nutritional importance. At the time of the Romans, for example, it was used in the training diet of gladiators, a term that in Latin means *Horderarii*, literally "men of barley." It has been the staple food especially in areas characterized by unfavorable environmental conditions and low productivity, as in the developing countries. It is still a staple food in some world areas, like North Africa and Near East regions and in the mountain region of Central Asia and South America. Something changed during the Middle Ages as wheat reached more significance in breadmaking, while barley almost disappeared from the diet of many countries, while its most important employment was for feed and, to a lesser extent, in alcohol beverages production (Giraldo et al. 2019). The cause can be found in the low gluten content, which determines less swelling of the loaf, and lower quality in general.

Barley is one of the cereals having higher genetic diversity. According to the sowing period, there are winter, spring, or alternative types. According to the morphology of the spike, there are two- and six-rowed types. The grain is generally hulled (covered or dressed) with an outer structure around the grain, but hull-less (naked) grains, in which the outer hulls detach easily, as in wheat, also exist. Based on starch composition of the grain, normal, *waxy,* or high amylose barley type can be distinguished; high lysin, high β-glucan, and proanthocyanin-free varieties are also available. Two-rowed barley is the wild one that is grown in Europe, whereas the six-rowed is developed by gene mutation and has a triple crop yield. Both two-rowed and six-rowed barley genotypes have winter and spring types. Winter type is cultivated in fall (especially in Europe, the United States, and Canada), whereas spring cultivar is cultivated in spring or summer in the Mediterranean region (Farag et al. 2022).

Most of the cultivated barley, about 65%, is employed for animal feed, about 33% is destined for malting production, while only 2% is used for human consumption as an ingredient of several food products (Farag et al. 2022). Historically, barley has been an important food source in many countries, including the Middle East, North Africa, and Northern and Eastern Europe (Farag et al. 2022). However, better product quality of food products prepared from wheat and rice compared to barley considerably decreased the use of barley as food, especially in the nineteenth and twentieth centuries. In the past, barley has received more attention from breeders and researchers thanks to the growing requests of healthy and more traditional food, which has led to the development of breeding programs aimed to obtain new barley varieties with a high level of healthy substances in the grain.

The growing world population is causing a higher demand for food and puts high pressure on soil and production systems. The future of agriculture will have to take into account the improvement of the nutritional quality of its products, trying to mitigate the environmental impact.

1.2 An Overview of Barley Composition

Barley grain consists of about 65–68% starch, 10–17% protein, 4–9% β-glucan, 2–3% free lipids, and 1.5–2.5% minerals. β-Glucans, part of soluble fiber, constitute about 75% of the barley endosperm cell walls together with 20% arabinoxylans and protein. Genotype, environment, and crop management practices used during the crop-growing cycle determine grain composition and structure. β-Glucans, which represent normally about 5–10% of the caryopsis weight, have a very significant effect on technological and quality properties of barley. For example, both β-glucans and arabinoxylans are fundamental in malting quality as they are responsible for wort viscosity and beer filtration rates and block hydrolytic enzymes, which in turn break down starch and protein within the cell walls. Moreover, for malting quality and brewer production, the beer foam stability is linked to a minimal amount of β-glucans in the malt (Farag et al. 2022). Accordingly, low β-glucan content of grain and/or its breakdown during malting are critical issues in brewing and most barley breeding has produced barley cultivars that are normally low in β-glucans.

Chemical and nutritional compounds of barley caryopsis can significantly vary due to genotype, agronomical practices, and environmental condition. The most abundant compounds of barley kernel are starch, fiber (including β-glucans), and proteins: varying the content of one of these substances will have a direct effect on the other two. Starch content is inversely correlated to protein and fiber. The *starch* is localized in the endosperm layer and is quite variable in composition, representing 54–75% of the grain weight. The starch of the normal or nonwaxy barleys has relatively high amylose (25%) and low amylopectin (75%), whereas waxy barleys have a low concentration of amylose (0–5%) and high amylopectin (100–95%). There are also high amylose barley mutants (up to 45% amylose fraction) (Farag et al. 2022).

The *nonstarch polysaccharides* of barley caryopsis are structural molecules of the cell walls in hull, aleurone, and endosperm tissue. These polysaccharides are included in the total dietary fiber. In contrast to starch, they are not digested by human digestive system, but they are fermented by intestinal microbiota, producing several breakdown products, including short-chain fatty acids. The major nonstarch polysaccharides in barley are β-glucans, arabinoxylans, and cellulose. Cellulose is mainly located in the hull of the grain.

Arabinoxylans are located mainly in the aleurone, where they consist about 71% of the fiber content; arabinoxylans are also present in the endosperm, where they account for 21% of the fiber content. Generally, arabinoxylan ranged from 3.8% to 6.0% based on genetic and environmental factors. Six-rowed genotypes contained higher levels of arabinoxylan than two-rowed varieties.

β-Glucans are structural cell component that forms the cell wall in barley endosperm that accounts for 75% of the endosperm cell wall mass. The level of β-glucans in barley has a significant genetic variation, ranging from 3.0% to 15–17% (Farag et al. 2022). *Waxy* barley showed higher concentration of β-glucans. β-Glucan content is also affected by environmental factors: for example, the concentration increases under dry and hot conditions, while decreases under moist conditions.

The level of *protein* in barley is highly variable: it is reported to vary from 8% to 30% as a percentage of total mass (Jaeger et al. 2021), although the level in typical barley is more commonly reported between 9% and 13%.

The grain protein content is a valuable quality factor, defining barley grain end-use value: high protein grains would be desirable for feed, while a lower level (between 10% and 12%) is desirable for malting (Jaeger et al. 2021). The protein content is highly variable and can be affected by genotype, environmental conditions, and the use of fertilizers.

Hordeins are the most abundant protein fraction: they belong to the prolamin class, namely storage proteins, characterized by a high content of glutamine and proline. These protein types are complex polymorphic mixtures of polypeptides. The other barley proteins are represented by a mixture of albumins, globulins, and glutenin (Jaeger et al. 2021). Prolamins are composed of high content of the amino acids glutamine, proline, and low level of essential amino acids such as lysine, threonine, and tryptophan.

After hordeins, glutelin are the second most abundant storage proteins: they contain high levels of glutamine, proline, and glycine, and other hydrophobic amino acids (Jaeger et al. 2021).

Albumin and globulin are minor protein forms in the aleurone grain and embryo (Farag et al. 2022).

The protein in barely includes both essential (about 28%: leucine, valine, threonine, phenylalanine, isoleucine, lysine, histidine, methionine) and nonessential amino acids (72%: glutamic acid, aspartic acid, proline, tyrosine, glycine, and cytosine), potentially making barley a good source of protein in food supplements. Among the essential amino acids, barley is lacking in lysine (Farag et al. 2022).

1.3 Importance of Barley in the Prevention of Chronic Diseases

There is growing evidence that higher intakes of whole grains are associated with reduced incidence and mortality from several chronic diseases.

Reynolds et al. (2019), in a recent systematic review and meta-analyses, showed that the data for whole grain consumption, from prospective studied and clinical trials, exhibit a reduction in all-cause mortality, coronary heart disease, cancer death, incidence of type 2 diabetes, and stroke mortality. The observed reductions in risk were high, typically around 20% with significant dose–response relationships. Among the bioactive components of whole grains, total fiber tended to explain the association to lower risk of CVD-related outcomes.

Barley is a cereal with a high level of dietary fiber, particularly the soluble fiber β-glucans, which is mainly credited for barley's health benefits, such as a reduction in cholesterol and glycemic response, the modulation of gut microbiota, the management of blood pressure, and a reduction in the incidence of metabolic syndrome (Murphy et al. 2020). In the scientific literature, there is a lot of evidence of the positive role of these compounds on physiological effects and the correlation between the consumption of β-glucans and blood cholesterol levels is now clarified. Several mechanisms have been proposed to elucidate the health effects of β-glucans and are still under study, but the factor thought to be the most responsible is the capability of soluble β-glucans to form viscous solution in the stomach and intestinal tract. The viscosity of β-glucan is related to its molecular weight, size, and its concentration in solution (Bai et al. 2019).

Cultivar, growing conditions, processing, and food matrix affect the physicochemical and health properties, particularly molecular weight and solubility, of β-glucans. For these reasons, when clinical trials and new "functional foods" are developed, these aspects should be strongly considered.

With consumers' increasing awareness of the importance of nutrition and health, food is not only regarded as a medium for satiety but also as a means for disease prevention and control; such desire has created a rapidly increasing demand for "functional foods," defined as foods that have an additional physiological benefit besides providing basic nutritional needs and opened markets for a broad range of processed foods, including beverages with specific health attributes (Shvachko et al. 2021).

Barley can be considered an excellent cereal to produce functional foods, thanks above all to the presence of soluble fiber (β-glucans). The strong scientific evidence regarding the metabolic effects of the consumption of β-glucans has led the EFSA to authorize the use of specific nutritional and health claims, thus allowing and regulating the communication to the public of the benefits associated with the consumption of nutrients, provided they are present in the final product in adequate quantities and in a form that can be used by the body. Therefore, it has ordered the claim for barley and oat-based products so that it can be reported that consume 3 grams per day of β-glucans, or 0.75 g per serving in the four main meals, helps reduce plasma levels of total cholesterol and low-density lipoprotein (LDL) (ESFA 2011).

In the past, thanks to the reported physiological properties of β-glucans, researchers have extensively studied the possibility to incorporate β-glucans in several food types with the aim to develop functional foods. Barley fractions enriched in β-glucans can be incorporated into wheat blends to make bakery good with improved nutritional properties. Finocchiaro et al. (2012) used two β-glucan-enriched flour fractions, from two hull-less barley cultivars with normal starch type and a high-β-glucan *waxy* genotype, mixed with bread wheat flour. The two flour blends were used to make bread, and the postprandial glucose response was determined. The results showed that incorporation of high level of β-glucans (6% in the final blend) from the non*waxy* cultivar lower the GI significantly. Blandino et al. (2015) used pearling fractions obtained from naked barley to substitute conventional refined wheat flour for breadmaking. These fractions resulted in abundant insoluble

and soluble dietary fiber (particularly β-glucans) and other bioactive compounds. The various healthy components present in barley kernel are not homogeneously distributed, and sequential pearling has shown to be an effective process to obtain fractions enriched in bioactive compounds (Blandino et al. 2015). Martínez-Subirà et al. (2020) obtained biscuits rich in β-glucans and antioxidants using barley flour from a purple hull-less cultivar. A single biscuit gives more than 0.75 g of β-glucans. In this way, it is easy to reach the target of EFSA recommendation and obtain the label lettering of "reduces blood cholesterol and risk of heart disease" (Martínez-Subirà et al. 2020).

Incorporation of β-glucan into pasta products revealed a lower glycemic response. A study on spaghetti enriched with barley β-glucans showed a β-glucan dose-dependence reduction of GI, with a reduction of up to 54% of GI with the incorporation of 10% of barley β-glucans (Tosh and Bordenave 2021). Similarly, a reduced glycemic index was reported in β-glucan-enriched breakfast bars (Mejía et al. 2020).

Diets and foods rich in β-glucans and with low glycemic index may have an effect on the prevention of important disease as coronary heart diseases and diabetes (Murphy et al. 2020). The simplest and most frequent use of β-glucan is their incorporation in cereal-based products, but their inclusion in beverages and dairy products has been also evaluated; it can also find some applications in the production of low-fat ice creams and yogurts (Mejía et al. 2020).

2 Barley Bioactive Compounds

In barley grains, it is possible to find a lot of bioactive compounds: from carbohydrate polymers (like β-glucan, arabinoxylan, and hemicellulose), to vitamin E (tocopherol and tocotrienols), to secondary metabolites like phenols, folates, and lignans. The focus of this section is to describe the bioactive phytochemicals that most characterize barley grain for its effects on human health: **β-glucans**, **tocols**, and **phenol/polyphenol compounds**.

2.1 β-Glucans

2.1.1 β-Glucans: Structure and Contents

Mixed-linkage (1→3),(1→4) linear β-D-glucans (β-glucans) are major constituents of endosperm cell walls of cereals such as barley (Fig. 2), oat, and to a lesser extent wheat and rye.

In barley, β-glucans account for 75% of the total cell wall polysaccharides, the remaining portion is made up of arabinoxylans, cellulose, glucamannans, and proteins. A unique feature of barley grain is that the β-glucan is uniformly distributed in the endosperm while it is concentrated in aleurone layers in oat grain, thus pearling or removal of the outer layers does not significantly affect the β-glucan content in barley (Murphy et al. 2020).

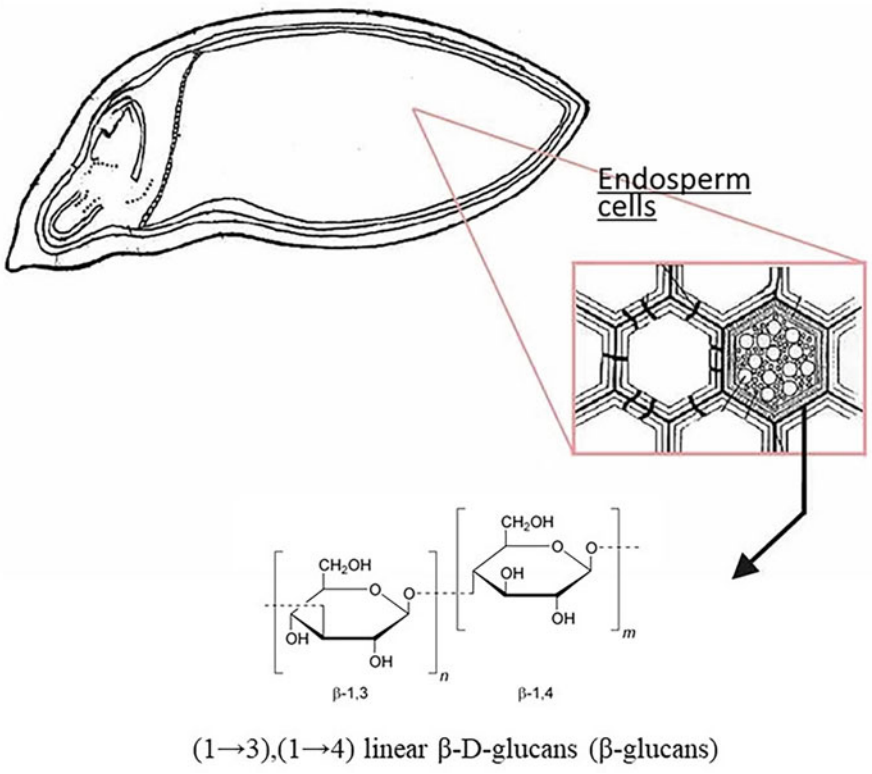


Fig. 2 Barley β-glucan structure and localization in the grain

β-Glucans are complex nonstarchy polysaccharides consisting of D-glucose monomers linked through β-glycosidic bonds, which widely exist in plants and microorganisms. Different sources of β-glucans lead to different structures and physicochemical properties (Bai et al. 2019).

Barley β-glucans are composed of linear homopolysaccharides of D-glucopyranosyl residues with β-(1→3) and β-(1→4) linkages. About 90% of glucose units are organized into blocks of cellotriose residues (three glucose molecules) and cellotetraose residues (four glucose molecules) joined by β-(1→3) bonds, while the rest are composed of segments longer cellulosics, composed of multiple glucose units (Bai et al. 2019).

As already discussed earlier, the molecular characteristics of β-glucans seem to determine their physical properties, such as water solubility, dispersibility, viscosity, and gelation properties, as well as of their physiological function. The chemical composition of β-glucans determines their partial solubility in water. For example, β-glucans containing blocks of adjacent β-(1→4) linkages can show lower solubility thanks to interchain aggregation (Bai et al. 2019). Therefore, in addition to the physiological benefits of soluble dietary fiber, cereal β-glucans also show health

properties typical of insoluble fiber, such as increased fecal bulk to reduce constipation and improve weight loss.

Barley and oat are primary cereal sources for β-glucans. Total β-glucan content of barley grains ranges from 2.5% to 11.3% by kernel weight, but they most frequently fall between 4% and 7%. However, some genotypes demonstrate to accumulate up to 13–17% of total β-glucans (Farag et al. 2022).

The variances in barley β-glucan concentration have been attributed to environmental growing conditions, although the genotype is the main factor in determining the final concentration of β-glucans in barley kernel. Naked barleys often have higher β-glucan contents (Farag et al. 2022) and are mainly used as human food thanks to easier processing after harvesting and consumption. Barley genotypes show also differences in the amylose/amylopectin ratio and genotypes with both *waxy* starch (with up to 100% amylopectin) and high amylose (over 35%) are available. Interestingly, these genotypes with different starch composition are also characterized by higher β-glucan concentration (Ferrari et al. 2009).

Barley generally contains high molecular weight β-glucans that determine high-viscosity solution and are responsible to enhance gut viscosity that is correlated to barley physiological activities (Murphy et al. 2020).

2.1.2 β-Glucans: Biochemical Pathway of Production

β-(1→3),(1→4) at 4–5 days after pollination (DAP) and are uniformly distributed in the tissue by 10 DAP. β-(1→3),(1→4)-glucan content increases considerably within 16–36 DAP, coinciding with the grain filling and maturation stages (Garcia-Gimenez et al. 2019).

The synthesis of β-glucans in barley is regulated by members of the cellulose synthase-like gene families, *CslF* and *CslH*. Seven *CslF* and one *CslH* (cellulose synthase-like genes) are responsible for β-glucan synthesis, and a UDP-glucose 4-epimerase (*UGE*) is also implicated. Among these, *HvCslF6* is the most highly transcribed and is the most important gene for β-glucan synthesis in barley kernel. *HvCslF6* was well characterized as a (1→3),(1→ 4)-β-glucan synthase and it is expressed throughout grain development. This was confirmed in cslf6 mutant lines (called beta-glucanless or *bgl*) that have very low levels of mixed linkage β-glucan. During grain development, *HvCslF6* and *HvCslF9* are the predominant expressed genes, and a mutation in the *HvCslF6* locus led to a loss of β-glucans accumulation. Burton et al. (2011) succeeded in overexpressing *HvCslF4* and *HvCslF6* in transgenic with extra copies of these genes, and the final result was an increase in grain β-glucan content of more than 50% and 80, respectively (Garcia-Gimenez et al. 2019).

β-Glucan concentration is a quantitative trait, and there are several associated QTLs: for example, one QTL is located on chromosome 7H, within 5 cM of the *Nud* gene; this suggests a possible linkage effect. Many QTLs involved in β-glucan accumulation overlap with QTLs for amylose content probably due to co-localization or interaction between genes (Meints et al. 2021).

In addition, several researchers suggested that variation in barley (1→3), (1→4)-β-glucan levels was also affected by polysaccharide remodeling and degradation pathways. Two barley genes, *HvGlbI* and *HvGlbII*, encoding for (1→3),

(1→4)-β-glucan end hydrolases, have been reported to contribute to this trait, having an impact on modification on β-glucans during the malting process (Garcia-Gimenez et al. 2019).

2.1.3 β-glucans: Physiological Properties and Functions in Relation to Human Health

β-Glucans play important roles, particularly in healthy foods and pharmaceutical products, due to their widely known beneficial effects, including immunomodulation, antitumor activity, serum cholesterol and glucose reduction, and obesity prevention (Bai et al. 2019). The degree and type of biological activity appear to be strictly correlated to β-glucans structural properties.

Physiological Properties of β-Glucans: Glycemic Control

The reduction of the glycemic response after consumption of food (glycemic control) is among the most intensely studied and well-documented properties of barley β-glucans. This effect has been investigated in numerous animal and human studies using model food systems.

There is evidence that increasing molar weight and/or viscosity of β-glucans improves the postprandial glucose regulation, indicating that this effect is based on the formation of a high-viscosity gel in the intestine. A proposed mechanism for these observations is that high viscosity of the intestinal content may slow down digestion of starch (by decelerating the diffusion of α-amylase toward its starch substrate) and the absorption of glucose (Tosh and Bordenave 2021).

β-Glucan molecular weight should be an important trait in glycemic response, but to date, there are still contradicting results regarding the importance of this trait on glycemic response. Some researchers found that high molar mass and viscosity are essential parameters for healthy effects, such as glycemic control (Schmidt 2022). In contrast, recent studies reported no differences in the control of blood glucose and cholesterol lowering between hydrolyzed and native barley β-glucan in mice (Schmidt 2022).

Physiological Properties of β-Glucans: Cholesterol Lowering

Another well-studied health-promoting property of barley β-glucans is the ability to lower blood cholesterol levels. While high-density lipoprotein (HDL) cholesterol remains unaffected by β-glucans, the reduction in low-density lipoprotein (LDL) cholesterol, an important marker for the risk of cardiovascular diseases, is recognized (Schmidt 2022). The generally accepted mechanism behind this is based on the formation of a highly viscous gel inside the human intestine. The gel interacts with the bile acids present in the gastric system, preventing their readsorption and thus promotes the de novo synthesis of such acids. Since cholesterol is used for the bile acid synthesis, it is captured from blood and its concentration decreases (Schmidt 2022).

Based on this data, the EFSA and FDA have concluded that the regular consumption of at least 3 g of β-glucans from oat and/or barley per day is required to achieve a substantial cholesterol-lowering effect (EFSA Panel on Dietetic Products, Nutrition and Allergies (NDA) 2011; FDA_b-glucan health 2006).

Physiological Properties of β-Glucans: Effects on Gut Microbiota

The role of microbiota in maintaining good health is now clearly established, and whole-grain barley can support the growth and maintenance of gut microbiota (Tosh and Bordenave 2021). Fermentable fiber and β-glucan are a fundamental part of it and could actively impact the microbiota.

β-Glucans perform their function of regulating blood cholesterol levels through the modulation of the microbiota, which is directly related to the metabolism of bile acids. β-Glucans are resistant to digestion, reach the small intestine, and are fermented by gut microbiota producing short-chain fatty acid (SCFA), limiting HMG-CoA activity, improving cholesterol catabolism (Murphy et al. 2020).

Through dietary fiber fermentation, intestinal microbiota produces short-chain fatty acids (SCFA) (acetic, propionic, and butyric acids): butyric acid, in particular, plays an important role in promoting and maintaining colon health, while propionic acid, after supplementation of β-glucans, showed hypocholesterolemic properties. Furthermore, SCFA have been found to modulate the immune system (Murphy et al. 2020).

β-Glucans are considered prebiotic as they improve indirectly gastrointestinal function by enhancing the intestinal microbiota (Murphy et al. 2020). They elicit the growth and activity of commensal bacteria in the colon. Both in vitro and in vivo models have confirmed that the growth of normal intestinal bacteria (*Lactobacilli* and *Bifidobacteria* species) is increased by β-glucans (Murphy et al. 2020).

The intestinal microorganism can also influence hormone secretion. In particular, insulin sensitivity may be improved by the promotion of gut hormone secretion from enteroendocrine cells by SCFAs (Murphy et al. 2020).

2.1.4 Methods of Nutraceutical Improvement: Agronomic and Postharvesting Techniques

β-Glucan content is a trait highly influenced by genotype, and this is a good point to develop new genotypes with the aim of functional food production. The environment and agronomic practices can also impact β-glucan levels. Supplementation of soil nitrogen can cause an increase in β-glucan levels, while increased irrigation has shown to lower β-glucan content (Choi et al. 2020). Dry and hot weather can also determine an increase in β-glucan concentration, while cooler, wetter conditions are generally correlated to a reduction in β-glucan content (Meints et al. 2021).

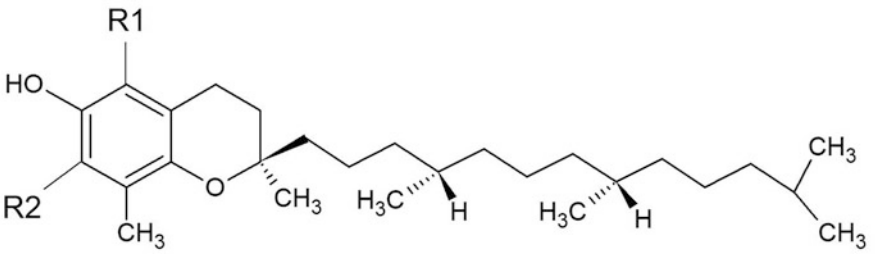
Given the nutraceutical importance of β-glucans, over the years, methods have been developed to incorporate barley flours into baked products to make them functional foods. As β-glucans are water-soluble compounds, barley fractions enriched in β-glucans were obtained by extraction with aqueous solvents (Mejía et al. 2020).

However, these procedures are highly energy-consuming and sometimes use dangerous solutions like sodium hydroxide or organic solvents, residues of which are forbidden in human foods. Otherwise, a way to increase β-glucan content is to mill and sieve barley flours to discriminate β-glucan-rich fraction (Mejía et al. 2020). The addition of barley flour, however, significantly worsens the quality characteristics of bakery products.

2.2 Tocols

2.2.1 Tocols: Structure and Content

Tocopherols (vitamin E) and tocotrienols, grouped as tocols, are a class of lipid-soluble antioxidants only synthesized by plants and other photosynthetic organisms. They are amphipathic molecules with a polar head group, the chroman ring, and a hydrophobic tail, which is either saturated (tocopherols) or contains three double bonds at positions 3, 7, and 11 (tocotrienols). Four homologs (α-, β-, γ-, δ-) exist for each tocol class, differing only in the number and position of methyl substitutions on the aromatic ring (Tiwari and Cummins 2009) (Fig. 3).

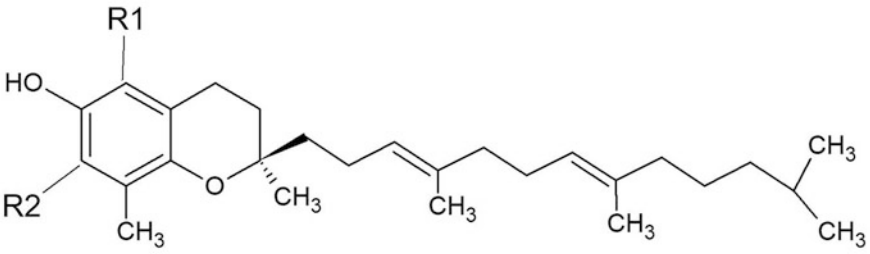


α-tocopherol R1= R2= CH_3

β-tocopherol R1= CH_3, R2= H

γ-tocopherol R1= H, R2= CH_3

δ-tocopherol R1= R2= H



α-tocotrienol R1= R2= CH_3

β-tocotrienol R1= CH_3, R2= H

γ-tocotrienol R1= H, R2= CH_3

δ-tocotrienol R1= CH_3, R2= H

Fig. 3 Structure of barley tocols

Oils from vegetable sources are the main forms of tocols, but also cereals, like barley, oats, wheat, rye, and rice are reported to be an excellent source of this class of compounds. The tocol composition of barley is unique as it includes all eight homologs (Obadi et al. 2021).

In the barley grain, tocopherols and β-tocotrienol are mainly located in the germ, while hulls and endosperm have substantial tocotrienols concentration (Obadi et al. 2021) and accumulate during grain development (Obadi et al. 2021). Among cereals, barley is one of the best sources of tocopherols and tocotrienols due to a high level and favorable distribution of all eight biologically active homologs (Obadi et al. 2021). In whole-grain barley, α-tocotrienol is the most individual tocol homolog, contributing about 47.7% of the total tocol content, followed by α-tocopherol, γ-tocotrienol, γ-tocopherol, β-tocotrienol, and δ-tocotrienol (Idehen et al. 2017). The total tocotrienol level is markedly higher than the total tocopherol level (Idehen et al. 2017).

There is a wide variation, influenced by genotypes and environment, in quantities of tocols in barley. Temelli et al. (2013) examined the total tocol content of whole-grain barley varieties, which ranged from 40 mg/kg to 151.1 mg/kg. Hulled barley has been shown to possess more tocols than hull-less barley (Idehen et al. 2017). Cavallero et al. (2004) also reported higher tocol and α-tocotrienol contents in hulled barley (53 and 61 mg/kg) than in hull-less barley (50.9 and 53.1 mg/kg). This was explained by the presence of tocols in the hull of barley.

2.2.2 Tocols: Physiological Properties and Functions in Relation to Human Health

The presence of the phenolic hydroxyl group in tocopherol and tocotrienols is essential for the antioxidant activity of vitamin E due to the ability to donate a phenolic hydroxyl group of the chromanol ring to blocking peroxidation of lipids in cell membranes (Tiwari and Cummins 2009). α-Tocopherol content is important from a nutritional point of view as it is the tocol homolog with the highest vitamin E activity. The other homologs of tocols exhibit vitamin E activity in the following order: α-tocopherol > β-tocopherol > α-tocotrienol > γ-tocopherol > β-tocotrienol> δ-tocopherol (Gangopadhyay et al. 2015). Although α-tocopherol has the highest vitamin E activity, α-tocotrienol has been found to possess 40–60 times higher antioxidant activity than α-tocopherol (Gangopadhyay et al. 2015).

In addition to antioxidant properties, studies have shown that tocotrienols have several beneficial functions. For example, thanks to the capacity of inhibition of cholesterol biosynthesis, they determine a protective effect by lowering LDL cholesterol (Tiwari and Cummins 2009). The hypocholesterolemic effect of α-tocotrienol is due to the suppression of hydroxyl-β-methylglutaryl co-enzyme A reductase, the key enzyme of cholesterol synthesis (Tiwari and Cummins 2009). Moreover, the tocol content of cereals can confer human health benefits, including modulating degenerative diseases like cancer and cardiovascular diseases (CVD) (Obadi et al. 2021).

2.2.3 Tocol Biosynthesis

The biosynthesis of tocopherol occurs in plastids of plant cell, except for the first step of tocopherol biosynthesis (biosynthesis of chromanol head) that occurs in a plant's cytoplasm (Fritsche et al. 2017).

The biosynthesis starts through the formation of the chromanol ring, which is derived via the shikimate pathway from homogentisate. The polyprenyl side chain (phytyl diphosphate, phytyl PP/PDP) is suggested to originate from DOXP (1-deoxy-D-xylulose-5-phosphate) pathway, as well as from the recycling of free phytol derived from the chlorophyll degradation process (Fritsche et al. 2017). In tocopherols, the prenyl tail attached to the chromanol ring is derived from phytyl diphosphate (PDP). In tocotrienols, it originates from the condensation of HGA and phytyl diphosphate (PDP), which is catalyzed by the plastid-localized enzyme HGA phytyl transferase (HPT) (Yang et al. 2011). Plant tissues vary enormously in their tocol content and composition with photosynthetic tissues generally containing low levels of total tocols and a high percentage of α-tocopherol, whereas seeds contain 10–20 times this level of total tocols, but α-tocopherol is often a minor component (DellaPenna and Pogson 2006).

2.2.4 Methods of Nutraceutical Improvement: Agronomic and Postharvesting Technique

Food processing significantly influences the level and stability of tocols (Tiwari and Cummins 2009). After harvesting, cereal grains undergo substantial changes in composition before being consumed; for example, tocol cereal content is influenced by pearling, milling, extrusion, cooking, malting, and baking (Tiwari and Cummins 2009).

Pearling (which causes the removal of hull, aleurone, and germ) significantly decreased tocol concentration, if compared to the whole seed, suggesting that dehulled seed is not a rich source of tocols, whereas its by-product (the removed material during pearling) is rich in tocols (Farag et al. 2022). Tocol level was not affected by malting process, although brewers' spent grains were enriched in tocols (Peterson 1994), becoming a good source of tocols. This attribute could make them valuable additions to food products, particularly as they will contain substances that have favorable effects on cholesterol levels and may offer remarkable economic and waste management opportunities.

Storage is another factor that has an influence on the tocol content. Antioxidants and, among them, the tocols can be easily affected by light, water, and heat. Generally, after being harvested, barley will be stored for periods ranging from 4 to 18 months before processing. The storage may have an effect on vitamin E content and antioxidant capacity. Do et al. (2015) measured tocol content and antioxidant capacity in 25 barley cultivars before and after 4 months storage at 10 °C. They observed that six homologs (α- T, α-T3, β-T, β-T3, γ-T, and γ-T3) were detected in barley genotypes after 4 months storage, but the amounts of δ-T and δ-T3 were minimal. The total vitamin E content changed from 6% to 30% after 4 months of storage.

Liu and Moreau (2008), on the other hand, found the storage at 35 °C with 75% relative humidity for 3 weeks caused no change in oil and tocopherols, but significant changes in tocotrienols. The changes were differential among T3 homologs, with α-tocotrienol decreasing and δ-tocotrienol increasing. The reduction in α-tocotrienol was accelerated in fractions characterized by a higher proportion of endosperm tissue. When the storage was prolonged for 6 months, the authors

confirmed these findings. These results and practical information can help us to produce barley fractions enriched with functional lipids and maintain stability of their products.

2.3 Simple Phenols and Polyphenols

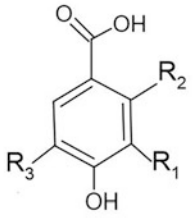
Phenolic compounds comprise a large heterogeneous class of secondary metabolites. Phenolics possess one or more aromatic rings with one or more hydroxyl groups and can be classified into different ways, based on their chemical structure, and distribution in nature.

Barley is considered a good source of many classes of phenolic compounds, such as phenolic acids, proanthocyanins, quinones, flavanols, chalcones, flavones, and flavonones. These compounds may exist in free, conjugated, or bound forms. Due to the high level of phenolics in barley grains, this cereal can be considered a good dietary source of antioxidants, which have antiradical and antiproliferative capacity and are correlated to the potential prevention of several chronic disease and health well-being (Farag et al. 2022).

2.3.1 Phenolics Acids: Structure and Content

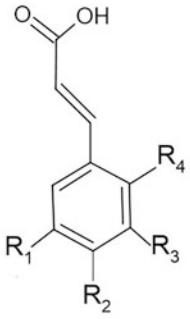
Benzoic and cinnamic acid derivatives (grouped as phenolic acids, Fig. 4) are the dominant class of phenolics in barley and are in the outer layers of the kernel (Idehen et al. 2017).

In barley, they are mainly present in the bound form, some are present in conjugated form, and only to a lesser extent are present in the free form. The free



Benzoic acids

p-Hydroxybenzoic acid	R_1= R_2= R_3= H
2,4-Dihydrixybenzoic acid	R_1= R_3= H, R_2= OH
Vanillic acid	R_1= OCH_3, R_2= R_3= H
Syringic acid	R_1= R_3= OCH_3, R_2= H



Cinnamic acids

o-cumaric acid	R_1= R_2= R_3= H, R_4= OH
p-cumaric acid	R_1= R_2= R_4= H, R_2= OH
Ferulic acid	R_1=OCH_3, R_2= OH, R_3=R_4= H
Sinapic acid	R_1= R_3= OCH_3, R_2= OH, R_4= H

Fig. 4 Structure of barley phenolic acids

forms are concentrated in the outer layers of the pericarp, whereas the bound forms are esterified to cell wall constituents, such as lignin, cellulose, arabinoxylans, and other polysaccharides (Loskutov and Khlestkina 2021). Ferulic and *p*-coumaric acids are phenolic compounds abundantly found in whole grains of barley; they exist in free, bound, and soluble-conjugated forms. The composition of phenolic compounds depends upon the grain type, cultivars, and morphological fraction used for analysis.

Deng et al. (2020) analyzed different hull-less barley genotypes for phenolic composition and content: bond ferulic acid resulted the predominant one, and the content varies from 50.16 to 54.65 mg/kg. In the study conducted by Šimić et al. (2019), ferulic acid content was found to have a wide range (4.5–102.7 mg/100 g dry weight) among different fractions of hull-less barley. They reported that ferulic acid was distributed mostly in the outer bran and has the lowest content in the endosperm. *p*-hydroxybenzoic (3.43–17.44 mg/kg), gallic (3.50–32.90 mg/kg), and vanillic (3.20–24.70 mg/kg) acids were the predominant phenolic acids present in the free form (Šimić et al. 2019).

Martinez et al. (2018) observed a considerable variation in the content of total ferulic acid concentration, analyzing a set of barley genotypes, hulled and hull-less, with different origin. The content of some bound phenolic acids was related to the occurrence or lack of hull, with significantly higher levels of these compounds observed in the hulled genotypes compared to the hull-less samples.

2.3.2 Flavonoids: Structure and Content

Flavonoids (Fig. 5) are a phenolic class with a C6–C3–C6 skeleton (two aromatic rings joined by a three-carbon link).

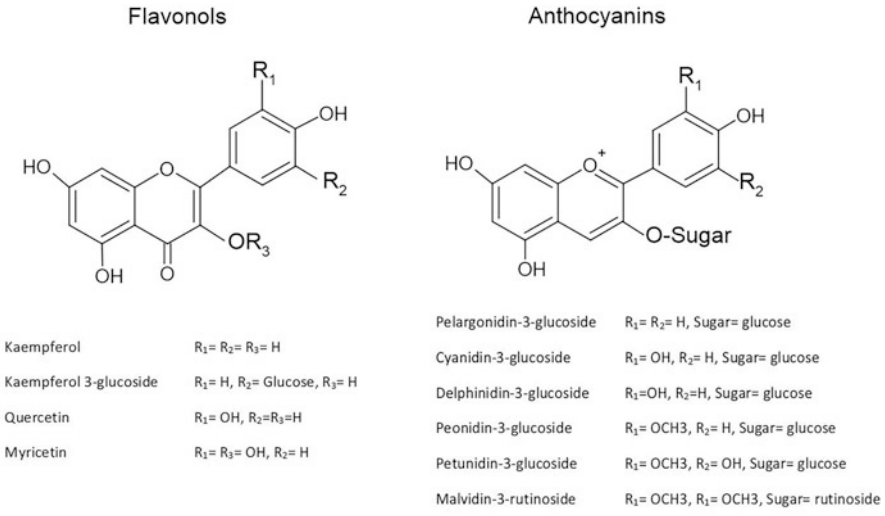


Fig. 5 Structure of barley flavonoids: flavonols and anthocyanins

Clinical trials indicate that flavonoids may be the phytochemicals present in cereal grains correlated to the moderation of many diseases, like cancer and coronary heart diseases (Idehen et al. 2017).

Flavonoids are a subclass of plant phenols and have been shown to have many health benefits, such as antioxidant, anticancer, antiallergic, and anti-inflammatory.

Naringin (flavanone), catechin (flavonol), and quercetin (flavon) were reported to be the three predominant flavonoids in hull-less barleys with contents varied from 0.8 to 9.8, 0.1–20.5, and 1.4–8.7 mg/100 g (dry weight), respectively (Deng et al. 2020).

Zhu et al. (2015) reported that total flavonoids (free and bound) in four hull-less barley varieties ranged between 145.5 and 247.4 mg catechin equivalents/100 g, dry weight basis, observing that it was higher than other grains like corn, wheat, oat, and rice (Zhu et al. 2015).

Anthocyanins are water-soluble flavonoids, located in the vacuole, and they are mainly present in the pericarp or the aleurone layers of barley kernel, determining the purple or blue grain pigmentation, respectively. This class of flavonoid is normally present as glycoside in the grains. A range of anthocyanin compounds have been identified in barley kernels by various authors. Martinez et al. (2018) identified up to 27 anthocyanins in different barley genotypes. The most common anthocyanin in purple barley is cyanidin 3-glucosode, followed by peonidin 3-glucoside and pelargonidin 3-glucoside. Delphinidin 3-glucoside was found to be the main anthocyanin in blue and black barley varieties (Martínez et al. 2018) (Fig. 5).

Martinez et al. (2018) found that purple and blue barleys contained higher average level of anthocyanins than the black barleys tested. This could be explained because the black pigmentation present in barley is linked to the presence of melanin-like pigment. So, it could be assumed that the color of the black barley genotypes is probably a result of co-pigmentation between anthocyanins and melanin-like pigments (Glagoleva et al. 2022).

Proanthocyanidins (PAs), or condensed tannins, are a group of polyphenols that naturally occur in a wide range of plants, including barley (Fig. 6).

PAs can be structurally distinguished by the different hydroxylation patterns of the basic flavan-3-ol units and by the nature of the bond among monomers. The common basic units are (epi)gallocatechins, (epi)catechins, and (epi)catechins, resulting in the formation of prodelphinidin, procyanidin, and propelargonidin structures, respectively. The units are commonly linked through B-type bonds (C4→C6or C4→C8 linkage), while an additional linkage (C2→O5 or C2→O7) contributes to the formation of A-type bonds (Zhu 2019).

Different proanthocyanidin composition and concentration has been observed and linked to genetic diversity in barleys. For example, Verardo et al. (2015) found that PA concentrations in 14 barley genotypes ranged from 293 to 653 mg/kg. The HPLC-MS analysis revealed the presence of catechin/epicatechin monomers, procyanidin dimer, prodelphinidin dimer, procyanidin trimer (36.7–167 mg/g), prodelphinidin trimer I (monogalloylated), prodelphinidin trimer II (digalloylated), procyanidin tetramer, prodelphinidin tetramer, and procyanidin pentamer (Verardo et al. 2015).

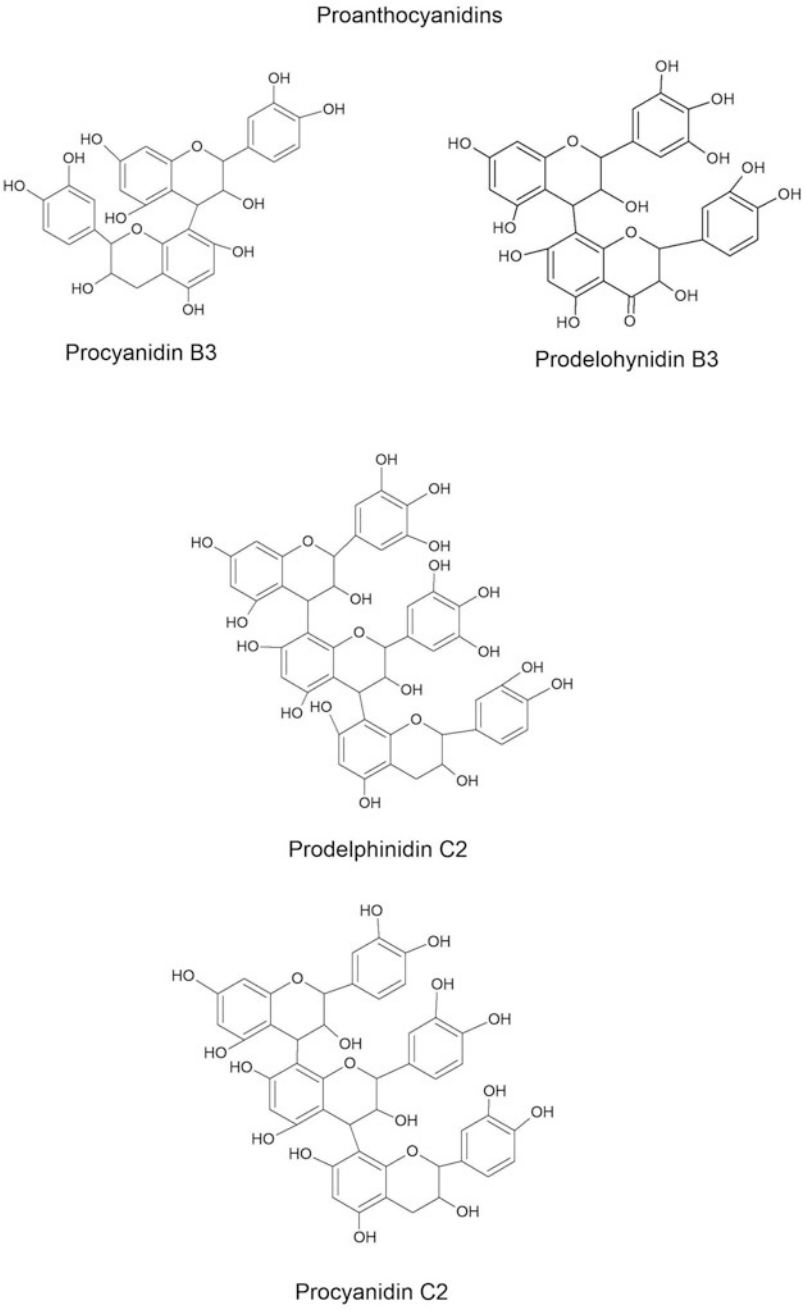


Fig. 6 Structure of proanthocyanidins found in barley

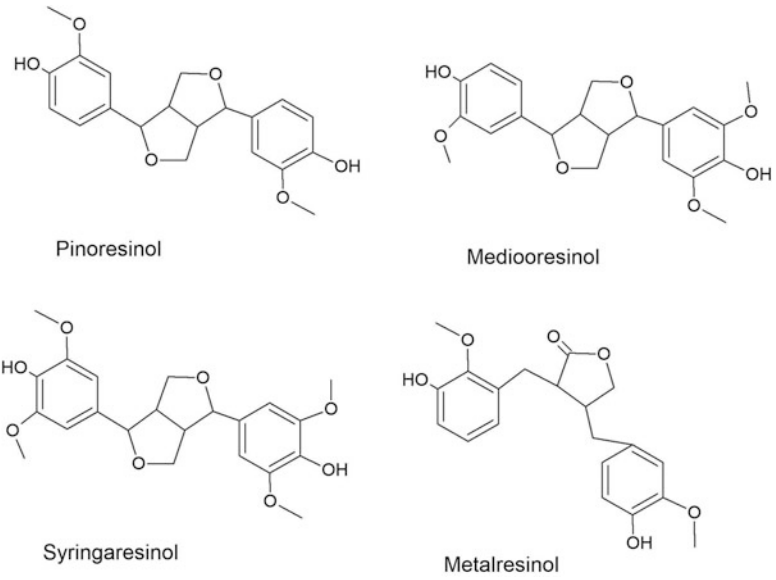


Fig. 7 Structures of major lignans in barley

Lignans (Fig. 7) are natural polyphenols widely distributed in the plant kingdom and are recognized as natural defense compounds.

They possess a phytoestrogens-like activity due to their structural and functional similarity to 17-β-estradiol. Lignans have been suggested to have a wide range of biological effects, such as antioxidant, antitumor, antimicrobial, antifungal, estrogenic, and antiestrogenic activities, and protect against coronary heart diseases. There is little information in the literature about the structure and concentration of lignans in barley. Smeds et al. (2007) analyzed lignan content in barley: the authors reported the presence of pinoresinol (71 mg/100 g), medioresinol (22 mg/100 g), syringaresinol (140 mg/100 g), lariciresinol (133 mg/100 g), cyclolariciresinol (28 mg/100 g), secoisolariciresinol (42 mg/100 g), secoisolariciresinol-sesquilignan (24 mg/100 g), matairesinol (42 mg/100 g), oxomatairesinol (28 mg/100 g), and 7-hydroxymatairesinol (541 mg/100 g) as major lignans and todolactol (127 mg/100 g), α-conidendrin acid (33 mg/100 g), nortrachelogenin (15 mg/100 g), and lariciresinol-sesquillgnan (6.6 mg/100 g) as minor lignans.

2.3.3 Simple Phenols and Polyphenols: Physiological Properties and Functions in Relation to Human Health

Free radicals in human body can induce a series of diseases such as cancer diabetes, atherosclerosis, and reproductive endocrine dysfunction. Phenols, in this sense, can have an important function as they act as scavengers of free radicals and could play a major role in moderating cardiovascular diseases. Several in vitro experiments

evidenced the bioactivities of barley phenols (Idehen et al. 2017). The most studied and reported bioactivity are the in vitro antioxidant capacity and free-radical-scavenging activity determined by different chemical tests, including DPPH, ORAC, FRAP, and ABTS (Obadi et al. 2021). Blue hull-less barley varieties evidenced higher antioxidant activities than regular barley varieties: these genotypes showed higher antioxidant than Canadian, Egyptian, and Tunisian barley (Yang et al. 2018). Li et al. (2019) demonstrated strong correlations between antioxidant capacities of barley grains and phenolic concentration, denoting that the phenolic compounds represent the main responsible for the antioxidant activity of whole naked barley flours. Abdel-Aal et al. 2012 selected Canadian and Egyptian barleys and investigated phenolic acid composition and antioxidant capacity against DPPH and ABTS radicals, and inhibition of oxidation of human low-density lipoprotein (LDL) cholesterol of whole grain flours and pearling fractions. The data showed significant variations among barley wholegrain flour and pearling/milling fractions in terms of phenolic acid composition and antioxidant capacity.

Health benefits of barley phenols have also been reported in in vivo experiments using animal and clinical models. Mice that were fed with 600 mg/kg BW of polyphenol extract from black highland barley showed a significant reduction in total cholesterol (23.33%), LDL cholesterol (26.29%), and atherosclerosis index (38.70%), and an increase in high-density lipoprotein cholesterol (HDL, 17.80%) (Shen et al. 2016). Lee et al. (2015) also demonstrated that extracts from barley sprouts containing polyphenols regulated AMP-activated protein kinase, a cellular sensor of energy metabolism and a regulator for cholesterol metabolism. Oxidative stress, which has a key role in pathology like diabetes and obesity, can be reduced by phytochemicals present in barley, particularly phenolic compounds (Idehen et al. 2017).

As previously reported, a large part of phenol compounds is bonded to cell wall constituents (i.e., cellulose, lignin, etc.), and therefore they are a part of dietary fiber. There is evidence that dietary fiber provides various beneficial health effects, among which are the decrease of the risk of cardiovascular disease, metabolic syndrome, diabetes, obesity, and cancers (Murphy et al. 2020); these health properties may be attributed to phytochemicals bound to or trapped within dietary fiber, especially phenolic compounds (Tosh and Bordenave 2021).

2.3.4 Simple Phenols and Polyphenols: Biosynthesis

Barley grains accumulate a large amount and type of phenolic compounds, which are a product of secondary metabolism in plants. Moreover, different barley genotypes present different grain coloration: the most studied are yellow, purple, red, blue, and black. The yellow color is linked to the presence of proanthocyanidins synthesized in seed coat of the kernel; purple and red pigmented genotypes are characterized by the presence of anthocyanins, synthesized in pericarp and glumes; blue color is caused by anthocyanins synthesized in aleurone layer of the grain. In white barley grains, the pigments are not present (Shoeva et al. 2016).

In barley, phenol biosynthesis metabolic pathway is well characterized: structural genes encoding enzymes of the pathway, as well as regulatory genes, have been described (Deng and Lu 2017) (Fig. 8).

The fists step in phenylpropanoid biosynthetic pathway is the phenylalanine synthesis from the shikimate pathway. Phenylalanine ammonia lyase (PAL) catalyzes this first enzymatic step, which cause the deamination of phenylalanine to generate cinnamic acid, which in turn is hydroxylated to *ρ*-coumaric acid, catalyzed by cinnamate 4-hydroxylase (*C4H*). 4-Coumaroyl CoA ligase (*4CL*) is involved in the catalysis of the formation of ρ-coumaroyl-CoA from ρ-coumaroyl acid.

The initial steps of the phenylpropanoid pathway are jointly known as the general phenylpropanoid pathway (GPP) (Deng and Lu 2017). After GPP, the pathway is branched leading to the synthesis of flavonoids, stilbenes, monolignols, phenolic acids, and coumarins. In the barley genome, the Chalcone synthase (CHS) characterize a gene family with at least seven copies. One of the *Chs* gene copies has been mapped to the short arm of chromosome 1H. Three nonoverlapping genetic markers for the *Chs* gene have been mapped to chromosomes 1HS, 1HL, and 6HS (Shoeva et al. 2016). The gene related to chalcone flavanone isomerase gene (*Chi*) has been identified and mapped to the long arm of chromosome 5H. Flavanone 3-hydroxylase gene (*F3h*) has been also characterized, and the gene has been localized to chromosome 2HL. Barley dihydroflavanol reductase gene (*Dfr*) has been mapped on the long arm of chromosome 3H. Other important genes coding for key enzyme have

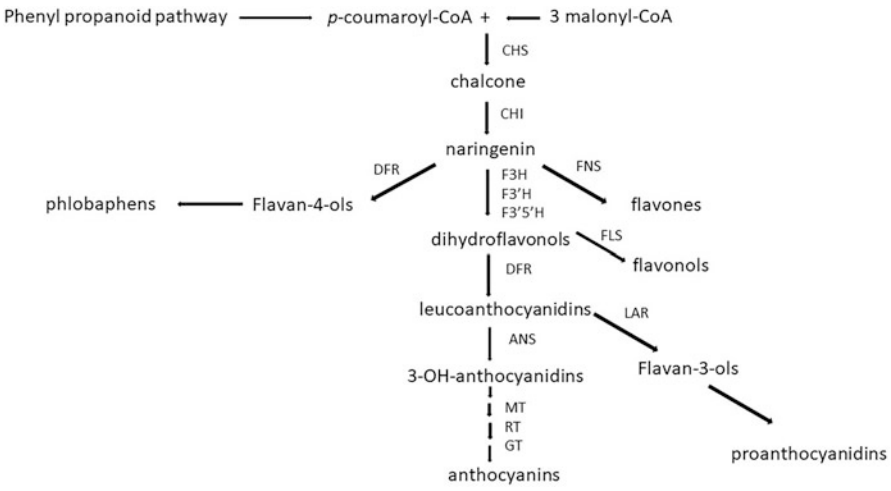


Fig. 8 A schematic presentation of the flavonoid biosynthetic pathway in barley. The enzymes reported are CHS (chalcone synthase); CHI (chalcone-flavanone isomerase); F3H (flavanone 3-hydroxylase); FLS (flavonol synthase); FNS (flavone synthase); F3′H (flavonoid 3′-hydroxylase); F3′5′H (flavonoid 3′,5′-hydroxylase); DFR (dihydroflavonol 4-reductase); ANS (anthocyanidin synthase); GT (glycosyltransferase); MT (methyltransferase), RT (rhamnosyltransferase); and LAR (leucoanthocyanidin reductase)

been identified in barley: barley leucoanthocyanidin reductase (LAR) and gene for UDP glucose-flavonol 3-O-glucosyltransferase (UFGT), which mapping on the short arm of chromosome 7H. Flavonoid 3'-hydroxylase (*F3'h*) and anthocyanidin synthase (*Ans*) genes were identified and localized on chromosomes 1H and 5HL, respectively (Shoeva et al. 2016).

In barley, anthocyanin pigmentation of the kernel pericarp is controlled by two complementary genes, *Ant1* (MYB) and *Ant2* (bHLH), which are located on chromosomes 7H and 2H, respectively (Glagoleva et al. 2020). They form, jointly with the transcription factor WD40, the MBW complex, which controls the anthocyanins biosynthetic pathway in a tissue-specific manner. Accumulation of blue anthocyanins in barley aleurone is controlled by *HvMyc2* gene, a paralog of *Ant2* (Strygina et al. 2017).

Melanins are the product of enzymatic oxidation of phenolic precursors, such as tyrosine, cinnamic acid derivatives, and catechol, to quinones, which then undergo subsequent polymerization. Melanin biosynthesis in barley is under the monogenic control of *Blp1* located on chromosome 1H (Long et al. 2019). Purple pigmentation (characterized by anthocyanins) and black pigmentation (characterized by phytomelanins) are under different genetic control, so these pigments can be both accumulated in the kernel combining genes controlling the two traits (Glagoleva et al. 2022).

2.3.5 Methods of Nutraceutical Improvement: Postharvesting Techniques

Many authors have investigated the effects of grain type, genotype, and environment effect on the accumulation of phytochemical in barley grain. The most important goal of these studies is the possibility to select barley varieties, with improved nutritional characteristics. Several studies reported that pigmented barley genotypes accumulate a high concentration of different classes of polyphenols, like anthocyanins in blue or purple barley. This class of flavonoid had demonstrated important healthy properties. Moreover, there is a large variation in composition among the phytochemical composition in different cultivars, reflecting factors related to genetic diversity, indicating the need for proper selection of genetic material (Irakli et al. 2020).

One of the most effective methods to improve nutritional value of seeds is germination. Several authors have demonstrated that germination can enhance the content of many phytochemicals, like phenolic compounds and other bioactive substances in grains, as well as remove antinutritional factors such as enzyme inhibitors. Tang et al. (2021) investigated the effect of germination on the polyphenol content, antioxidant capacity, and physicochemical properties of three different pigmented cultivars of hull-less barley. They observed that the germination determined an increase in the total phenolic and flavonoid content, as well as antioxidant properties. This could be due to the improved biosynthesis of polyphenols during germination.

Also, malting process significantly affects the polyphenol content in grains. During the malting process, barley grains undergo different biochemical and physical changes: hydrolytic enzymes are released in the shoot or root developed during the malting process are removed physically (Sharma et al. 2022). Moreira et al. (2013) studied the antioxidant properties of brewer's spent grain and reported that malt produced at lower temperature (light malt) had the higher total phenol content than raw barley. Fogarsi et al. (2015) observed that barley total phenols significantly increased after malting process. The germination and malting process increased the phenolic contents in barley due to the release of bound phenolic content by inherent enzymes (Sharma et al. 2022).

Pearling, an important primary process in food-barley utilization, is a milling approach that gradually removes grain tissue by abrasive action and that, consequently, modulates the distribution of nutrients and phytochemicals in barley flour. Irakli et al. (2020) evidenced that bioactive components, including phenolic acids and flavonoids, exhibit a decreasing concentration from the external layers to the center part of barley grain among all barley cultivars tested, in contrast to what happens for β-glucans.

3 Next-Generation Breeding for Phytochemicals and Nutrient Contents

In recent years, scientists and the food industry developed and studied a wide range of new cereal products designed with healthy properties. Cereals, and among them barley, are a very rich source of several nutrients, micronutrients, and phytochemicals as already discussed in Sect. 2. Breeding techniques can help us to obtain new genetic materials with high a concentration of these cereal phytochemicals, and at the same time help to develop high-yielding genotypes, being able to combine the maximum level achievable for nutritional compounds and high-quality indicators, like yield, disease resistance, etc. (Loskutov and Khlestkina 2021).

The genetic diversity existing among genotypes is at the base of all crop improvement programs. Diversity can be described as the degree of differentiation between or within species. The existence of genetic diversity within and between crop species permits the breeders to select superior genotypes to be used as parent in breeding programs. Wild species, related species, mutant lines, etc., represent the important source of genetic diversity and may provide new and favorable alleles (Bhandari et al. 2017).

Thanks to the genetic diversity existing between the breeding parents, it is possible to obtain heterosis and transgressive segregants. Genetic diversity helps breeders to develop varieties for specific traits like quality improvement and tolerance to biotic and abiotic stresses. Another very important aspect, especially with respect to the continuous and recent climatic changes, is that genetic diversity is an aspect that helps crop plants to adapt to different environments and to variations in the environment itself (Bhandari et al. 2017).

3.1 Tools for Assessing Genetic Diversity in Genomic Era

As genetic diversity is a fundamental aspect in breeding, the assessment of this diversity is crucial to ensure sufficient further genetic gain and make informed selection decisions. DNA-based methods are the most used for genetic diversity assessments. They can be divided into fragment analysis-based methods, in which DNA polymorphisms are determined by different sizes of DNA fragments; hybridization array-based methods, where DNA polymorphisms are detected by hybridizing the sample DNA to arrayed probes; and sequencing-based methods, where sequencing is used to detect polymorphisms. With these DNA-based methods, even thousands of polymorphisms may be assessed simultaneously in a lot of sample plants, making the analysis more cost-effective. Some of these can detect both coding and noncoding regions of the genome, which provides a broad view of genetic diversity and could potentially be employed to find genomic loci under selection. In recent years, DNA sequencing technologies have been subjected to considerable development, also allowing the analysis of complex genomes and the study of their genetic diversity (Loera-Sánchez et al. 2019).

Thus, DNA-based molecular markers started to be used to examine the genetic basis of agronomic traits and improve crops through genome-based methods, including marker-assisted selection (MAS) and, in recent years, genomic selection. One of the disadvantages of traditional phenotype-based selection is the time needed to develop new varieties. Compared with traditional phenotype-based selection, genome-based breeding can directly find favorable alleles underlying the desired traits, and thus leads to precise selection that significantly reduces the time needed to develop new varieties (Liu et al. 2021).

The fundamental characteristics of molecular markers are that they are correlated with phenotypic expression of a genomic trait, and that they are stable and detectable in all tissues despite of plant growth stage, and differentiation and status of the cell; moreover, environmental, pleiotropic, and epistatic effects do not influence them.

3.2 Connecting Genotype to Phenotype

3.2.1 Marker-Assisted Selection (MAS)

As mentioned above, the development of molecular marker technology made plant breeding became more efficient by means of marker-assisted selection (MAS).

Molecular markers are sequences of nucleotides and can be explored through the polymorphisms present between the nucleotide sequences of a given population. The polymorphism is highlighted through the identification of deletions, insertions, gene mutation, duplication, and translocation of a precise sequence, and they do not really influence the function of genes. There are five important considerations for the application of DNA markers in MAS: reliability (markers must be firmly connected to the target loci); quality and nature of required DNA; amount of DNA required; methodology for marker examination (high-throughput, straightforward techniques

are required); cost-effectiveness; the level of polymorphism (ideally, the marker should be highly polymorphic in the breeding material): the suitability of a given molecular marker is dependent on its capacity to identify polymorphisms in the nucleotide sequences permitting segregation between the different alleles (Hasan et al. 2021).

Different types of molecular markers have been used over the last few decades: restriction fragment length polymorphism (RFLP), random amplification of polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), microsatellite or simple sequence repeat (SSR), sequence characterized amplified region (SCARs), cleaved amplified polymorphic sequences (CAPS), single-nucleotide polymorphism (SNP), and diversity arrays technology (DArT) markers. Among these, SNP markers were the most widely used (Hasan et al. 2021).

Molecular markers are powerful genomic tools for the assessment of genetic diversity within and between populations and offer the possibility to improve the performance of plant breeding program, especially in QTL identification, which help us to identify and locate genes on chromosomes. Moreover, if the identified marker is closely related to important agronomic traits, it can be used in MAS and thus increase the efficiency of selection, allowing acceleration of the total breeding process (Al-Abdallat et al. 2017).

Marked-assisted selection combines information derived from mapping position of agronomical importance traits with and the linked molecular. The success of MAS depends on several factors, such as the number of individuals that can be analyzed, the genetic nature of the trait and the background in which the target gene must be transferred, and it is achieving increasing importance in breeding programs.

Regarding the implications of barley breeding for nutrition quality traits, recent findings of regulatory features of anthocyanin biosynthesis in barley were useful for both MAS approach and genetic editing-based breeding strategies.

Strygina et al. (2017) identified and characterized components of the anthocyanin synthesis regulatory network in the aleurone layer of barley. The genes identified and characterized included elements of the regulatory complex MBW, from which *HvMyc2*aMYC-encoding gene appeared to be the main factor determining variation of barley aleurone pigmentation.

Gordeeva et al. (2019) applied a marker-assisted backcrossing approach evidencing its efficacy in creating barley genotype with favorable alleles of anthocyanin regulatory genes: they developed a set of near isogenic lines (NILs) and revealed specific features of the anthocyanin biosynthesis regulation in barley pericarp: the dominant alleles of both the *Ant1* and *Ant2* genes were required for anthocyanin accumulation in pericarp. The dominant allele of the two genes was upregulated in purple-pigmented line. The activity of these genes also influenced the expression of the *F3'h* and *Ans* structural genes. In addition, positive effect between *Ant1* and *Ant2* was detected. The results achieved in this work represent a strong basis for target manipulation to modulate the content of anthocyanins in barley grains.

Since domestication, introgression breeding has been successfully used to improve barley, and it remains a key system for augmenting genetic diversity to deal with current and future challenges to crop production (Hernandez et al. 2020). Introgression essentially implies the transfer of a particularly and desirable trait from

one plant species to another with the help of hybridization and frequent back-crossing. Introgression of favorable alleles using marker-assisted selection is now faster and more efficient due to the development of high-throughput tools and technologies.

In the past century, breeding strategies were aimed to develop highly productive but uniform cultivars of cereals like wheat, barley, and other crops, leading to a constant decrease in genetic diversity. It is therefore necessary to find and improve strategies to increase the range of diversity especially in key quality traits. An important source of genes, alleles and genetic variation can be found in the ancestor of cultivated barley (*H. vulgare* subsp. *spontaneum*), landraces, and germplasm collections. Landraces have been used for the introgression of allele into adapted genotypes. Germplasm collections are excellent sources of genetic diversity. There are several germplasm collections in different part of the world, for example, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK, Germany), the Okayama University Barley and Wild Plant Resource Center (Japan), the International Center for Agricultural Research in the Dry Area (ICARDA), and the United States Department of Agriculture National Small Grains Collection (USDA-NSGC, the United States). All the collections from these institutions are an excellent source of generic diversity and were well characterized for gene and alleles linked to disease resistance, abiotic stress resistance, and yield performances (Hemshrot et al. 2019; Visioni et al. 2013; Monteagudo et al. 2019).

3.2.2 QTLs

Genetic mapping of major genes and quantitative trait loci (QTLs) were widely used for many important agricultural traits and integrated with the conventional breeding process.

A genome-wide association study (GWAS) has the same approach as the QTL mapping study, but using natural populations, landraces, cultivar collections released over years, and genotypes with little or no pedigree information. It is used to reveal the relationship between markers and phenotypic traits based on linkage disequilibrium (LD), considering that associated markers resulting from GWAS could be used in new MAS programs. GWAS approaches were used to identify novel QTLs for traits with practical implications on barley agronomical performances, quality, and disease-related traits.

β-Glucan content is a quantitative trait with many associated QTLs found on chromosomes 1H, 2H, 5H, and 7H; they have been mapped in several populations, and candidate genes have been identified and validated. The QTL located on chromosome 7H and is within 5 cM of the *Nud* gene (Swanston and Middlefell-Williams 2012), indicating a possible linkage effect. A pleiotropic effect between β-glucan and the recessive allele at the Waxy (*WX*) locus has been demonstrated (Meints et al. 2021). The Waxy gene codes for a granule-bound starch synthase I (*GBSSI*) (Li et al. 2021). Li et al. (2021) carried out a genome-wide association study (GWAS) and found several new QTLs that were related to starch content; these candidate genes and alleles can be used in future breeding programs focusing on amylose and amylopectin content. It has been also shown that mutations at the *Lys3*

and *Lys5* loci can modify β-glucan content, as well as other healthy compounds. Despite the shriveled endosperm, genotypes containing these mutations revealed higher amounts of β-glucans, fructans, arabinoxylans, and resistant starch (Meints et al. 2021).

In the study of Steele et al. (2013), a breeding program for nutraceutical compounds in barley was developed using a large segregating population, developed from crosses between naked and hulled parents, selecting naked barley that were more than 65% genetically similar to UK hulled barley. With another approach, the authors obtained lines from a mapping population, with β-glucan contents ranging from 1.4% to 8.6% with transgressive segregation occurring in both directions; they also evidenced no significant association of β-glucan content with the nude gene (*nud*).

A QTL describing about 30% of the β-glucan inheritance was found on chromosome arm 7HL, approximately 15 cM above *nud*, and *CslF6* is the most probable candidate gene (Steele et al. 2013).

If QTLs linked to β-glucan content and QTLS for yield components are combined, it is possible to develop new higher yielding functional food barley.

The other quality barley trait of particular interest, related to the importance that antioxidant compounds have risen in recent years, is phenolic compound accumulation in the grain. To improve this quality trait, it is essential to identify genes and/or QTLs responsible for phenolic metabolism, but there are only a few studies on QTLs associated with phenolic compounds in barley. Han et al. (2018) determined total phenolic total flavonoid content and antioxidant capacity in grains of Tibetan barley, analyzing both wild and cultivated accessions. The results showed wild barley had higher content and broader genetic diversity of phenolic compounds than cultivated genotypes. The authors identified 20 unique QTLs associated with total phenolic compounds, total flavonoids, and antioxidant activity in Tibetan wild barley. Wild and cultivated barleys showed clearly different presence of QTLs linked to phenolic compounds, highlighting a significant genetic difference between the two different genetic materials and showing that Tibetan wild barley is suitable for barley breeding for phenolic compound content.

The flavonoid class of compounds that characterized the pigmentation of barley pericarp are anthocyanins (Zhu 2018). Zhang et al. (2017) studied the genetic control of major anthocyanin compounds, peonidin-3-glucoside, P3G, and cyanidin-3-glucoside, C3G, which characterized the purple pericarp of barley using a segregating population suitable for QTL mapping. Both anthocyanin compounds were linked to two loci, one located on chromosome arm 2HL and the other on 7HS. The two different anthocyanins appear to be controlled by the different interactions of the two loci. The impact of the 7HS locus on P3G and C3G was not so clear if before the effect of the 2HL locus was removed. The biosynthesis of peonidin-3-glucoside needs at least one copy of 2HL alleles. This does not seem to be the case for the accumulation of cyanidin-3 glucoside that was produced regardless of the allele combinations between loci. The inheritance of purple pigmentation of the barley grains showed a typical maternal effect. The different anthocyanins present in pigmented barleys, their different concentrations, and their different genetic control

suggest the development of breeding programs, and consequently food processing, directed to individual anthocyanins.

Mohammadi et al. (2014) conducted an association mapping study on physicochemical properties of barley that play a role in its nutritional quality, like total phenolics, amylose, and β-glucan content. They used 3.069 breeding lines, including two-row and six-row spring barley, from US breeding programs, and 2.041 SNP markers for association mapping. For total phenolics, they identified three significant regions on 3H, 4H, and 5H chromosomes. Two regions on 2H and 7H were associated with β-glucan. They also identified several markers associated with amylose content on chromosome 7H.

The future of breeding will be, at genomic level, the development of genotypes that contain the most favorable alleles for every gene for target compounds. Thanks to new genomic technologies, which allows the employment of genome-wide marker assays, accurate and high throughput, and in combination with new methods like gene editing and rapid generation turnover such as genomic selection (GS), have the capacity to accelerate the rate of genetic gains in crop breeding programs.

Traits linked to quality and nutritional value of grains are the result of human selection in the relatively short period of 10,000 years since domestication began compared to other more complex traits like grain yield. As a result, many of these traits are relatively simple and controlled by one or a few major genes that have been selected by humans (Kumar Id et al. 2019). If we can use genomics to find and characterize these genes selected by humans, we can improve breeding strategies to obtain high-quality grains for breeding for grain quality.

3.2.3 Omic Technologies for Functional Food

The application of *omics* technologies to crop breeding is strongly growing in interest. Foodomics include different *omic* technologies in relation to food and nutrition science, with the aim of improving human health and well-being. The *omic* topics comprise four major broad areas like genomics, transcriptomics, proteomics, and metabolomics.

Genomics includes the sequencing of whole genomes, assembly and annotation of the sequences, identification and development of molecular markers and quantitative trait loci (QTLs) for target traits, genomics-assisted breeding, genomic selection, etc. (Nayak et al. 2021), and all these tools can be used for the improvement of the nutritional and functional quality of cereal species. The characterization of key genes involved in defining grain quality traits can be achieved by conventional genetic mapping techniques, but direct genomics approaches are helpful for the purpose. For example, genome resequencing of a large number of genotypes that differ for a specific trait can allow association with the quality trait.

Transcriptomics refer to the study of how genes are expressed, changed, and interconnected in specific tissue at a particular stage. The differential gene expression can be studied and quantified by using different molecular biology techniques such as RNA sequencing, microarrays, Serial analysis of Gene Expression (SAGE), qRT-PCR, etc. While microarray, SAGE, and qRTPCR technologies are used for defined transcripts, the RNA sequencing has the advantage of high-throughput

sequencing and captures all expressed genes (Nayak et al. 2021). Analysis of differential expression between genotypes that differ for specific traits is also a step to gene identification. RNA-seq analysis is now a well-established tool for analysis of differences in the transcriptome, which represent the sum of all RNA transcripts of a specific organism. The RNA-seq analysis of developing and mature grains can reveal the genetic determinants of grain composition and, as a consequence, grain quality. Chen et al. (2014) made a de novo transcriptome analysis of developing barley grains using two Tibetan hull-less barley landraces, focusing the attention on gene expression levels related to the biosynthesis of storage components (starch, protein, and β-glucan); the temporal and spatial patterns of these genes were deduced from the transcriptome data of cultivated barley Morex. The results of this study showed how the genes related to important nutritional traits changed the expression during germination and developing of the seed. Moreover, the characterization of these genes provides resources to identify genes that can help in nutritional quality improvement of hull-less barley.

Proteomics can be effectively used to study protein structure, function, and interaction with other proteins or other bioactive compounds. Advanced techniques like matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) and liquid chromatography coupled to mass spectrometry (LC-MS) can detect and identify proteins differentially expressed (Nayak et al. 2021).

Metabolomics is a systematic study of all the metabolites of organism, tissue, cell, and can be considered as a phenotyping tool. A typical and practical application of this *omic* science concerns the identification and quantification of specific metabolites present in a sample. Metabolomics can be used for quantification of biologically active compounds, food fingerprinting, and food profiling. Techniques like gas chromatography coupled to mass spectrometry (GC-MS), liquid chromatography coupled to mass spectrometry (LC-MS), and nuclear magnetic resonance (NMR), have been largely used for the purpose of identification and quantification of metabolites (Nayak et al. 2021). Metabolomic analyses have been generally classified as targeted or untargeted. Targeted analyses focus on a specific group of compounds, requiring their identification and quantification. Targeted analyses are important for assessing the behavior of a certain class of metabolites in the sample under specific conditions. Moreover, targeted metabolomics usually requires higher level of purification and selective extraction protocols. In contrast, with untargeted metabolomics the attention is focused on the detection of as many groups of metabolites as possible to obtain patterns or fingerprints without necessarily identifying or quantifying all compounds. Untargeted analyses have been used for the fingerprints of biological events, such as plant diseases. The development of untargeted metabolic approaches was facilitated by the rapid advance in analytical techniques allowing the simultaneous detection of a wide range of compound classes. The techniques most commonly used for the characterization of compounds are mass spectrometry (MS) and NMR. Several *omics* technologies have been used for barley especially for the study of disease resistance and yield-related traits. However, there are only a few reports related to the functional compounds and quality properties in barley (Han et al. 2018; Chen et al. 2014).

Liang et al. (2022) employed an untargeted metabolomics approach using UPLC-MS/MS to investigate the metabolic differences of five hull-less barley with different grain colors. The authors annotated more than 600 metabolites, including flavonoids, amino acids, and phenolic acids. Colored hull-less barley cultivars were rich in flavonoids and possessed specific metabolite profiles that differ from the white hull-less barley, and these flavonoids also vary between the different colored cultivars. These results improve the understanding of the metabolic pathways and health value associated with different barley pericarp colors.

Plant metabolomics is a powerful tool to explore the metabolic and molecular regulatory mechanisms of plant growth, stress responses, and the improvement of crop productivity and quality. Thanks to next-generation sequencing technologies, metabolome-based GWAS (mGWAS) has been used to study genetic pathway that determines metabolic diversity and their associations with complex traits in plants (Scossa et al. 2016). Metabolomics combined with other *omics* could be a key issue of agronomic performance that was not resolved previously. Besides the chemical information, plant metabolomics can provide data on correlation among the different metabolites and agronomic important traits. Even more promising is the possibility of studying the relationship between metabolite modification and the resulting phenotype (Scossa et al. 2016; Zeng et al. 2020). This approach has accelerated to elucidate flavonoid pathways in cereals, including barley. Zeng et al. (2020) conducted comprehensive metabolic profiling and a metabolite-based genome-wide association study (mGWAS) in the grain and leaf of 196 hull-less and hulled barley accessions. They identified a total of 90 loci associated to metabolites from different branches of the phenylpropanoid pathway, which are also involved in UV-B protection; some alleles related to high-level metabolite trait were found to be significantly enriched in naked barley, suggesting co-selection of various phenylpropanoids. They also identified some genetic determinants regulating natural variation of phenylpropanoid content, including three novel proteins, a flavone C-pentosyltransferase, a tyramine hydroxycinnamoyl acyltransferase, and a MYB transcription factor.

Besides these *omics* approaches, genome-editing tools like RNAi, CRISPR/Cas9, TALENs, and ZFNs can be utilized to improve the crop plants. Use of computational and bioinformatics tools is essential and indispensable while using all these technologies (Nayak et al. 2021).

Genome editing (also called gene editing) is one of the most powerful tools to study the function of genes and an approach by which it is possible to obtain desirable traits in crops. It consists of cutting the genome with a nuclease and then introducing new mutations through DNA repair pathways. Three genome-editing systems, ZFN (zinc-finger nucleases), TALEN (transcription activator-like effector nucleases), and CRISPR-Cas (clustered regularly interspaced short palindromic repeats/CRISPR-associated protein), have been well developed and reported for plants (Nadakuduti and Enciso-Rodríguez 2021). CRISPR-Cas9 is, so far, the most promising and versatile genome-editing technology and has been applied in barley to manipulate grain quality traits (Garcia-Gimenez and Jobling 2022).

3.3 Genetic and Genomics Resources

Several genomic resources have been developed to study partial or total genomic sequences and the related gene function in barley. High-quality barley genetic maps have been developed, based on mutant phenotypes, and these results updated and complemented with genome-wide genetic maps generated from molecular markers.

The open accessibility of barley sequencing data allowed us to understand the genetic and regulatory functions of genes related to agronomically important phenotypes. The haploid genome of barley is about 5.3 Gb in size distributed across seven chromosomes. In 2012, the International Barley Genome Sequencing Consortium published the first reference genome derived from the cultivar Morex (Mayer et al. 2016); this reference genome has been then improved in sequence depth, genome assembly, and annotation (Monat et al. 2019; Mascher et al. 2017) and can be accessed at http://barleysequence.org/.

With the release of the barley genome sequence and the open availability of data, barley breeding is now in the "genomic" era, while barley research is in the postgenomic era. A number of genomic databases for barley have been produced, most of them freely available, and they are being used in different ways to identify or map the specific genes or genomic regions (Riaz et al. 2021) (Table 1).

4 Barley Gene Pools

Wide genetic variability can be found inside the three barley gene pools, characterized by different levels of fertility with cultivated barley (Wendler 2018). The primary gene pool includes barley itself (*Hordeum vulgare* spp. L), collecting landraces, obsolete, and modern varieties, and *H. spontaneum*. The accessions belonging to this gene pool can easily produce hybrids and are therefore the basis of barley improvement.

H. bulbosum is the only member within the secondary gene pool. A relatively large set of hybrids, substitution, and introgression lines between barley and *H. bulbosum* have been obtained and characterized via genotyping-by-sequencing and Exome Capture to unlock these genetic resources.

The tertiary gene pool, including more than 31 wild species, has severe barrier to fertility respect to cultivated barley and its role in breeding programs is limited. An exception is represented by Tritordeum, an amphiploid derived from the cross between the wild barley *Hordeum chilense* and durum wheat. Tritordeum, from a nutritional point of view, has some peculiarities in comparison with the parentals, that is, the high carotenoid content and very low gluten content. The success of this new species underlines the importance of genetic resources for the development of new innovative products for agriculture and industry, as reviewed by Avila et al. (2021).

The three gene pools are the building block to obtain the species pan-genome and genus super-pangenome. A key tool for the study of the barley pangenetic diversity

Table 1 Barley genomic resources (Riaz et al. 2021; Jayakodi et al. 2020; Tan et al. 2020; König et al. 2020)

Database/Website URL	Application	Tools
EnsemblPlants http://www.ensemblgenomes.org	Web database that acquires the genomic and proteomic data of different plant species, including barley	Ve!P BLAST
Gramene http://www.gramene.org	Provides an overview of comparative maps of cereals including available updated molecular markers and maps of barley	BLAST
Nord-Gen https://www.nordgen.org/en/	International database for genetic stock and mutant data collection in Nordic countries	
BARLEX http://barlex.barleysequence.org	Presents the first linearly ordered barley sequence; provides physical and genetic maps of molecular markers and genes using different version of assembly and gene set, with expression profiling data of 16 developmental stages, as well as exome capture data	BLAST
MorexGenes https://ics.hutton.ac.uk/morexGenes/	Offers access to gene expression levels from RNA-seq data of the barley cv. Morex, which are assembled from whole-genome shotgun sequences of Morex	BLAST
GrainGenes https://wheat.pw.usda.gov.GG3/barley_blvd	Genetic database primarily containing data on barley and wheat (genetic markers, gene expression, QTLs)	BLAST CMAP
HvGDB http://www.plantgdb.org/HvGDB/	Barley database provided by plant genome DataBase; it offers comparative genomics by using genomic data integration and analysis It contains advanced tools for comparative genomics, to analyze syntenic relationships among grass genomes	BLAST GeneSeqer GenomeThreader
Bex-DB https://barleyflc.dna.affrc.go.jp/bexdb/index.html	Database developed with the availability of full-length cDNA libraries of a two-rowed malting barley, Haruna Nijo	BLAST
Barley DB http://earth.nig.ac.jp/~dclust/cgi-bin/index.cgi?lang=en	Includes material on barley germplasms and genome resources, as well as BLAST and extra tools, which enables the creation of graphical figures of BLAST query results	CMAP BLAST BLASTscope GenomePaint
BarleyVarDB	Database that provides data related to barley's genomic variations in the form of three datasets – SNPs, InDels, and whole-genome sequences of wild and cultivated barley genomes	BLAST Primer3

(continued)

Table 1 (continued)

Database/Website URL	Application	Tools
HarvEST https://harvest.ucr.edu/	HarvEST originated as EST database-viewing software in support of gene function analyses and oligonucleotide design, then grew to support activities including microarray content design, SNP identification, genotyping platform design, comparative genomics, and the coupling of physical and genetic maps	BLAST
BRIDGE https://bridge.ipk-gatersleben.de/#start	Visual analytics web Tool for barley Genebank Genomics	

and its exploitation for crop improvement is in fact the genome sequence. However, a single reference genome cannot capture the full complement of barley sequence diversity. The genomes of 20 barley accessions belonging to primary gene pool, including landraces, varieties, and one *Hordeum spontaneum*, have been sequenced. From the assembly of multiple high-quality sequences, the so-called barley pan-genome has been obtained that can be considered a genomic infrastructure able to capture the full complement of sequence diversity of a crop species (Jayakodi et al. 2020). The use of structural variants individuated has been investigated in 300 GenBank accessions. The authors concluded: "*This first-generation barley pan-genome makes previously hidden genetic variation accessible to genetic studies and breeding.*" At an even higher level of complexity, it is assumed that the so-called super-pangenome can also be developed for crops, including barley. The super-pangenome is the assembly of the pangenomes of different species, able to capture the genomic diversity at the genus level. Pangenomics of wild relatives belonging to the different gene pools can complete gene repertoire of a genus (Khan et al. 2020).

5 Conclusion and Future Perspective

Providing safe, nutritious, and accessible food continues to be a major challenge for agriculture.

Barley is characterized by the presence of a large set of phytochemicals with the potential, and sometimes already demonstrated, impact on human health, so they are expected to play an ever-growing role in food industries, which are always looking for new healthy food for the general population. Consequently, plant breeders were driven to develop new breeding programs aimed to obtain cereal crop cultivars with higher contents of bioactive components in the grain (Loskutov and Khlestkina 2021).

The progresses in genomics provide new modern tools for make breeding programs more efficient. Especially, the assembly of the first barley reference genome offered great opportunities for the application of genomics in plant breeding. Genomic information in combination with sequencing, resequencing, and genotyping

datasets is being utilized to study important agricultural traits and their linked genes (Riaz et al. 2021). Moreover, the integration of genomics with other omics approaches in barley science aims to provide a nutritionally rich and safe cereal (Nayak et al. 2021). Another important aspect is that genomic tools will have a primary role to understand the impact of several environmental factors such as heat, drought, and plant diseases on crop quality, and this will become essential in adapting to a changing climate.

Acknowledgments The authors acknowledge the financial support of the Joint Call Cofund FACCE ERA-GAS (grant no. 696356), Connect Farms project Italian MIPAAF project ID 79.

References

Abdel-Aal ESM, Choo TM, Dhillon S, Rabalski I (2012) Free and bound phenolic acids and total Phenolics in black, blue, and yellow barley and their contribution to free radical scavenging capacity. Cereal Chem 89(4):198–204

Al-Abdallat AM, Karadsheh A, Hadadd NI, Akash MW, Ceccarelli S, Baum M et al (2017) Assessment of genetic diversity and yield performance in Jordanian barley (Hordeum vulgare L.) landraces grown under Rainfed conditions. BMC Plant Biol 17(1):191

Ávila CM, Rodríguez-Suárez C, Atienza SG (2021) Tritordeum: creating a new crop species-the successful use of plant genetic resources. Plants 10(5):1029. https://doi.org/10.3390/plants10051029

Bai J, Ren Y, Li Y, Fan M, Qian H, Wang L et al (2019) Physiological functionalities and mechanisms of β-glucans. Trends Food Sci Technol 88:57–66

Bhandari H, Nishant Bhanu A, Srivastava K, Singh M, Hemantaranjan A (2017) Assessment of genetic diversity in crop plants-an overview. Adv Plants Agric Res 7(3):279–286

Blandino M, Locatelli M, Gazzola A, Coïsson JD, Giacosa S, Travaglia F et al (2015) Hull-less barley pearling fractions: nutritional properties and their effect on the functional and technological quality in bread-making. J Cereal Sci 65:48–56

Burton RA, Collins HM, Kibble NAJ, Smith JA, Shirley NJ, Jobling SA et al (2011) Over-expression of specific HvCslF cellulose synthase-like genes in transgenic barley increases the levels of cell wall (1,3;1,4)-β-d-glucans and alters their fine structure. Plant Biotechnol J 9(2):117–135

Cavallero A, Gianinetti A, Finocchiaro F, Delogu G, Stanca AM (2004) Tocols in hull-less and hulled barley genotypes grown in contrasting environments. J Cereal Sci 39(2):175–180

Chen X, Long H, Gao P, Deng G, Pan Z, Liang J et al (2014) Transcriptome assembly and analysis of Tibetan hulless barley (Hordeum vulgare L. var. nudum) developing grains, with emphasis on quality properties. PLoS One 9(5):e98144

Choi H, Esser A, Murphy KM (2020) Genotype × environment interaction and stability of β-glucan content in barley in the Palouse region of eastern Washington. Crop Sci 60(5):2500–2510

DellaPenna D, Pogson BJ (2006) Vitamin synthesis in plants: tocopherols and carotenoids. Annu Rev Plant Biol 57:711–738

Deng Y, Lu S (2017) Biosynthesis and regulation of Phenylpropanoids in plants. CRC Crit Rev Plant Sci 36(4):257–290

Deng N, Zheng B, Li T, Liu RH (2020) Assessment of the phenolic profiles, hypoglycemic activity, and molecular mechanism of different highland barley (Hordeum vulgare L.) varieties. Int J Mol Sci 21(4):1175

Do TDT, Cozzolino D, Muhlhausler B, Box A, Able AJ (2015) Antioxidant capacity and vitamin E in barley: effect of genotype and storage. Food Chem 187:65–74

Farag MA, Xiao J, Abdallah HM (2022) Nutritional value of barley cereal and better opportunities for its processing as a value-added food: a comprehensive review. Crit Rev Food Sci Nutr 62: 1092–1104

Food and Drug Administration, HHS (2006) Food labeling: health claims; soluble dietary fiber from certain foods and coronary heart disease. Final rule Fed Regist 71(98):29248–50

Ferrari B, Finocchiaro F, Stanca AM, Gianinetti A (2009) Optimization of air classification for the production of β-glucan-enriched barley flours. J Cereal Sci 50(2):152–158

Finocchiaro F, Ferrari B, Gianinetti A, Scazzina F, Pellegrini N, Caramanico R et al (2012) Effects of barley β-glucan-enriched flour fractions on the glycaemic index of bread. Int J Food Sci Nutr 63(1):23–29

Fogarasi AL, Kun S, Tankó G, Stefanovits-Bányai É, Hegyesné-Vecseri B (2015) A comparative assessment of antioxidant properties, total phenolic content of einkorn, wheat, barley and their malts. Food Chem 167:1–6

Fritsche S, Wang X, Jung C (2017) Recent advances in our understanding of tocopherol biosynthesis in plants: an overview of key genes, functions, and breeding of vitamin E improved crops. Antioxidants 6:99

EFSA Panel on Dietetic Products, Nutrition and Allergies (NDA) (2011) Scientific Opinion on the substantiation of health claims related to beta-glucans from oats and barley and maintenance of normal blood LDL-cholesterol concentrations (ID 1236, 1299), increase in satiety leading to a reduction in energy intake (ID 851, 852), reduction of post-prandial glycaemic responses (ID 821, 824), and "digestive function" (ID 850) pursuant to Article 13(1) of Regulation (EC) No 1924/2006. EFSA J 9(6):2207

Gangopadhyay N, Hossain MB, Rai DK, Brunton NP (2015) A review of extraction and analysis of bioactives in oat and barley and scope for use of novel food processing technologies. Molecules 20:10884–10909

Garcia-Gimenez G, Jobling SA (2022) Gene editing for barley grain quality improvement. J Cereal Sci 103:103394

Garcia-Gimenez G, Russell J, Aubert MK, Fincher GB, Burton RA, Waugh R et al (2019) Barley grain (1,3;1,4)-β-glucan content: effects of transcript and sequence variation in genes encoding the corresponding synthase and endohydrolase enzymes. Sci Rep 9(1):17250

Giraldo P, Benavente E, Manzano-Agugliaro F, Gimenez E (2019) Worldwide research trends on wheat and barley: a bibliometric comparative analysis. Agronomy 9:352

Glagoleva AY, Shoeva OY, Khlestkina EK (2020) Melanin pigment in plants: current knowledge and future perspectives. Front Plant Sci 11:770

Glagoleva A, Kukoeva T, Mursalimov S, Khlestkina E, Shoeva O (2022) Effects of combining the genes controlling anthocyanin and melanin synthesis in the barley grain on pigment accumulation and plant development. Agronomy 12(1):112

Gordeeva EI, Glagoleva AY, Kukoeva TV, Khlestkina EK, Shoeva OY (2019) Purple-grained barley (Hordeum vulgare L.): marker-assisted development of NILs for investigating peculiarities of the anthocyanin biosynthesis regulatory network. BMC Plant Biol 19:52

Han Z, Zhang J, Cai S, Chen X, Quan X, Zhang G (2018) Association mapping for total polyphenol content, total flavonoid content and antioxidant activity in barley. BMC Genomics 19(1):81

Hasan N, Choudhary S, Naaz N, Sharma N, Laskar RA. (2021) Recent advancements in molecular marker-assisted selection and applications in plant breeding programmes. J Genet Eng Biotechnol 19(1):128. https://doi.org/10.1186/s43141-021-00231-1

Hemshrot A, Poets AM, Tyagi P, Lei L, Carter CK, Hirsch CN et al (2019) Development of a multiparent population for genetic mapping and allele discovery in six-row barley. Genetics 213 (2):595–613

Hernandez J, Meints B, Hayes P (2020) Introgression breeding in barley: perspectives and case studies. Front Plant Sci 11:761

Idehen E, Tang Y, Sang S (2017) Bioactive phytochemicals in barley. J Food Drug Anal 25:148–161

Irakli M, Lazaridou A, Mylonas I, Biliaderis CG (2020) Bioactive components and antioxidant activity distribution in pearling fractions of different Greek barley cultivars. Foods 9(6):783

Jaeger A, Zannini E, Sahin AW, Arendt EK (2021) Barley protein properties, extraction and applications, with a focus on brewers' spent grain protein. Foods 10(6):1389
Jayakodi M, Padmarasu S, Haberer G, Bonthala VS, Gundlach H, Monat C et al (2020) The barley pan-genome reveals the hidden legacy of mutation breeding. Nature 588(7837):284–289
Khan AW, Garg V, Roorkiwal M, Golicz AA, Edwards D, Varshney RK (2020) Super-Pangenome by integrating the wild side of a species for accelerated crop improvement. Trends Plant Sci 25 (2):148–158
König P, Beier S, Basterrechea M, Schüler D, Arend D, Mascher M et al (2020) BRIDGE – a visual analytics web tool for barley Genebank genomics. Front Plant Sci 11:701
Kumar Id A, Mantovani EE, Simsek S, Jain S, Elias EM, Mergoum M (2019) Genome wide genetic dissection of wheat quality and yield related traits and their relationship with grain shape and size traits in an elite × non-adapted bread wheat cross. PLoSONE 14(9):e0221826. https://doi.org/10.1371/journal.pone.0221826
Lee JH, Lee SY, Kim B, Seo WD, Jia Y, Wu C et al (2015) Barley sprout extract containing policosanols and polyphenols regulate AMPK, SREBP2 and ACAT2 activity and cholesterol and glucose metabolism in vitro and in vivo. Food Res Int 72:174–183
Li Q, Yang S, Li Y, Huang Y, Zhang J (2019) Antioxidant activity of free and hydrolyzed phenolic compounds in soluble and insoluble dietary fibres derived from hulless barley. LWT 111:534–540
Li M, Geng L, Xie S, Wu D, Ye L, Zhang G (2021) Genome-wide association study on total starch, amylose and amylopectin in barley grain reveals novel putative alleles. Int J Mol Sci 22(2):1–13
Liang L, Li W, Tian M, Pan J, Feng Z (2022) Metabolomic profiling of five hulless barley (Hordeum vulgare L.) with different: study on the metabolite difference of five hulless barley with different colors based on metabonomics approach. Genet Resour Crop Evol 69(5):1843–1853
Liu KS, Moreau RA (2008) Concentrations of functional lipids in abraded fractions of hulless barley and effect of storage. J Food Sci 73(7):C569–C576
Liu J, Fernie AR, Yan J (2021) Crop breeding – from experience-based selection to precision design. J Plant Physiol 256:15331
Loera-Sánchez M, Studer B, Kölliker R (2019) DNA-based assessment of genetic diversity in grassland plant species: challenges, approaches, and applications. Agronomy 9:881
Long Z, Jia Y, Tan C, Zhang XQ, Angessa T, Broughton S et al (2019) Genetic mapping and evolutionary analyses of the black grain trait in Barley. Front Plant Sci 9:1921
Loskutov IG, Khlestkina EK (2021) Wheat, barley, and oat breeding for health benefit components in grain. Plan Theory 10:1–18
Martínez M, Motilva MJ, López de las Hazas MC, Romero MP, Vaculova K, Ludwig IA (2018) Phytochemical composition and β-glucan content of barley genotypes from two different geographic origins for human health food production. Food Chem 245:61–70
Martínez-Subirà M, Romero MP, Puig E, Macià A, Romagosa I, Moralejo M (2020) Purple, high β-glucan, hulless barley as valuable ingredient for functional food. LWT 131:109582
Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T et al (2017) A chromosome conformation capture ordered sequence of the barley genome. Nature 544:427–433
Mayer K, Waugh R, Langridge P, Close T, Wise R, Graner A et al (2016) A physical, genetic and functional sequence assembly of the barley genome. Nature 491(7426):711–716
Meints B, Vallejos C, Hayes P (2021) Multi-use naked barley: a new frontier. J Cereal Sci 102: 103370
Mejía SMV, de Francisco A, Bohrer BM. (2020) A comprehensive review on cereal β-glucan: extraction, characterization, causes of degradation, and food application. Crit Rev Food Sci Nutr 60(21):3693–3704
Mohammadi M, Endelman JB, Nair S, Chao S, Jones SS, Muehlbauer GJ et al (2014) Association mapping of grain hardness, polyphenol oxidase, total phenolics, amylose content, and β-glucan in US barley breeding germplasm. Mol Breed 34(3):1229–1243
Monat C, Schreiber M, Stein N, Mascher M (2019) Prospects of pan-genomics in barley. Theor Appl Genet 132:785–796

Monteagudo A, Casas AM, Cantalapiedra CP, Contreras-Moreira B, Gracia MP, Igartua E (2019) Harnessing novel diversity from landraces to improve an elite barley variety. Front. Plant Sci 10:434

Moreira MM, Morais S, Carvalho DO, Barros AA, Delerue-Matos C, Guido LF (2013) Brewer's spent grain from different types of malt: evaluation of the antioxidant activity and identification of the major phenolic compounds. Food Res Int 54(1):382–388

Murphy EJ, Rezoagli E, Major I, Rowan NJ, Laffey JG (2020) β-glucan metabolic and immunomodulatory properties and potential for clinical application. J Fungi 6:1–33

Nadakuduti SS, Enciso-Rodríguez F (2021) Advances in genome editing with CRISPR systems and transformation Technologies for Plant DNA manipulation. Front Plant Sci 11:637159

Nayak SN, Aravind B, Malavalli SS, Sukanth BS, Poornima R, Bharati P et al (2021) Omics technologies to enhance plant based functional foods: an overview. Front Genet 12:742095

Obadi M, Sun J, Xu B (2021) Highland barley: chemical composition, bioactive compounds, health effects, and applications. Food Res Int 140:110065

Peterson DM (1994) Barley tocols: effects of milling, malting, and mashing. Cereal Chem 71(1): 42–44

Reynolds A, Mann J, Cummings J, Winter N, Mete E, Te Morenga L (2019) Carbohydrate quality and human health: a series of systematic reviews and meta-analyses. Lancet 393(10170): 434–445

Riaz A, Kanwal F, Börner A, Pillen K, Dai F, Alqudah AM (2021) Advances in genomics-based breeding of barley: molecular tools and genomic databases. Agronomy 11:894

Schmidt M (2022) Cereal beta-glucans: an underutilized health endorsing food ingredient. Crit Rev Food Sci Nutr 62:3281–3300

Scossa F, Brotman Y, de Abreu e Lima F, Willmitzer L, Nikoloski Z, Tohge T et al (2016) Genomics-based strategies for the use of natural variation in the improvement of crop metabolism. Plant Sci 242:47–64

Sharma P, Goudar G, Longvah T, Gour VS, Kothari SL, Wani IA (2022) Fate of polyphenols and antioxidant activity of barley during processing. Food Rev Intl 38:163–198

Shen Y, Zhang H, Cheng L, Wang L, Qian H, Qi X (2016) In vitro and in vivo antioxidant activity of polyphenols extracted from black highland barley. Food Chem 194:1003–1012

Shoeva OY, Mock HP, Kukoeva TV, Börner A, Khlestkina EK (2016) Regulation of the flavonoid biosynthesis pathway genes in purple and black grains of Hordeum vulgare. PLoS One 11(10): e0163782

Shvachko NA, Loskutov IG, Semilet TV, Popov VS, Kovaleva ON, Konarev AV (2021) Molecules bioactive components in oat and barley grain as a promising breeding trend for functional food production. Molecules 26(8):2260

Šimić G, Horvat D, Lalić A, Komlenić DK, Abičić I, Zdunić Z (2019) Distribution of β-glucan, phenolic acids, and proteins as functional phytonutrients of hull-less barley grain. Foods 8(12):680.

Smeds AI, Eklund PC, Sjöholm RE, Willför SM, Nishibe S, Deyama T et al (2007) Quantification of a broad spectrum of Lignans in cereals, oilseeds, and nuts. J Agric Food Chem 55(4): 1337–1346

Steele K, Dickin E, Keerio MD, Samad S, Kambona C, Brook R et al (2013) Breeding low-glycemic index barley for functional food. Field Crops Res 154:31–39

Strygina KV, Börner A, Khlestkina EK (2017) Identification and characterization of regulatory network components for anthocyanin synthesis in barley aleurone. BMC Plant Biol 17:184

Swanston JS, Middlefell-Williams JE (2012) The influence of steep regime and germination period on the malting properties of some hull-less barley lines. J Inst: Brew 18:186–191. https://onlinelibrary.wiley.com/doi/10.1002/jib.21

Tan C, Chapman B, Wang P, Zhang Q, Zhou G, Zhang XQ et al (2020) BarleyVarDB: a database of barley genomic variation. Database 2020:baaa091

Tang Y, Xiao L, Wu X, Li W, Wu T, Zhang P (2021) Impact of germination pretreatment on the polyphenol profile, antioxidant activities, and physicochemical properties of three color cultivars of highland barley. J Cereal Sci 97:103152

Temelli F, Stobbe K, Rezaei K, Vasanthan T (2013) Tocol composition and supercritical carbon dioxide extraction of lipids from barley pearling flour. J Food Sci 78(11):C1643–C1650

Tiwari U, Cummins E (2009) Nutritional importance and effect of processing on tocols in cereals. Trends Food Sci Technol 20:511–520

Tosh SM, Bordenave N (2021) Emerging science on benefits of whole grain oat and barley and their soluble dietary fibers for heart health, glycemic response, and gut microbiota. Nutr Rev 78:13–20

Verardo V, Cevoli C, Pasini F, Gómez-Caravaca AM, Marconi E, Fabbri A et al (2015) Analysis of oligomer proanthocyanidins in different barley genotypes using high-performance liquid chromatography-fluorescence detection-mass spectrometry and near-infrared methodologies. J Agric Food Chem 63(16):4130–4137

Visioni A, Tondelli A, Francia E, Pswarayi A, Malosetti M, Russell J et al (2013) Genome-wide association mapping of frost tolerance in barley (Hordeum vulgare L.) BMC Genomics 14:424

Wendler N (2018) The genomes of the secondary and tertiary gene pools of barley. In: Stein, N., Muehlbauer, G. (eds) The Barley Genome. Compendium of Plant Genomes. Springer, Cham

Yang W, Cahoon RE, Hunter SC, Zhang C, Han J, Borgschulte T et al (2011) Vitamin E biosynthesis: functional characterization of the monocot homogentisate geranylgeranyl transferase. Plant J 65(2):206–217

Yang XJ, Dang B, Fan MT (2018) Free and bound phenolic compound content and antioxidant activity of different cultivated blue highland barley varieties from the Qinghai-Tibet plateau. Molecules 23(4):879

Zeng X, Yuan H, Dong X, Peng M, Jing X, Xu Q et al (2020) Genome-wide dissection of co-selected UV-B responsive pathways in the UV-B adaptation of Qingke. Mol Plant 13(1): 112–127

Zhang XW, Jiang QT, Wei YM, Liu C (2017) Inheritance analysis and mapping of quantitative trait loci (QTL) controlling individual anthocyanin compounds in purple barley (Hordeum vulgare L.) grains. PLoS One 12(8):e0183704

Zhu F (2018) Anthocyanins in cereals: composition and health effects. Food Res Int 109:232–249

Zhu F (2019) Proanthocyanidins in cereals and pseudocereals. Crit Rev Food Sci Nutr 59:1521–1533

Zhu Y, Li T, Fu X, Abbasi AM, Zheng B, Liu RH (2015) Phenolics content, antioxidant and antiproliferative activities of dehulled highland barley (Hordeum vulgare L.). J Funct Foods 19: 439–450

Oats: Nutritional Uniqueness and Breeding of a Healthy Superfood

Caterina Morcia, Franca Finocchiaro, Stefano Delbono, Roberta Ghizzoni, Fabio Reggiani, Paola Carnevali, Giorgio Tumino, Ilaria Carrara, and Valeria Terzi

Contents

C. Morcia · F. Finocchiaro · S. Delbono · R. Ghizzoni · F. Reggiani · V. Terzi (✉)
Council for Agricultural Research and Economics (CREA), Research centre for Genomics & Bioinformatics, Fiorenzuolad'Arda, PC, Italy
e-mail: valeria.terzi@crea.gov.it

P. Carnevali
Barilla S.p.A, Parma, PR, Italy

G. Tumino
Plant Breeding, Wageningen University and Research, Wageningen, The Netherlands

I. Carrara
Dipartimento di Scienze degli Alimenti e del Farmaco, Università degli Studi di Parma, Parma, PR, Italy

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_5

Abstract

Oat, of *Poaceae* grass family, is an important multipurpose cereal, cultivated for grain, food, feed, fodder, and straw, alone, in mixture, or as a dual-purpose crop. This cereal ranks sixth in world production statistics, following wheat, maize, rice, barley, and sorghum. Recently, an increased interest in oat arose due to its unique health-related properties. This cereal produces valuable and unique macro-, micro-, and phytonutrients, is rich in soluble fibers (mainly β-glucan), polyphenols, galactolipids, and contains a relatively high quantity of protein. In addition, oat is an important source of phenolic acids being the only cereal that contains avenanthramides, a group of phenolic alkaloids with beneficial effects on cardiovascular diseases and colon cancer prevention. Moreover, oat bran is an excellent source of B complex vitamins (B1, B2, B3, and B6) and tocopherol (vitamin E). The lipidic amount of whole-grain oats is doubled compared with other cereals and mainly consists of unsaturated fatty acids such as linoleic, oleic, and linolenic acids. Considerable genetic resources exist with large collections of landraces, wild relatives, and cultivars; therefore, sources of genetic variation are available for breeding purposes. The traditional breeding has been directed to improve yield and agronomical traits, such as abiotic and biotic stress resistance, lodging resistance, growth habit, together with a few quality-related traits, namely protein and starch content. During the last years, oats turned out to be a very interesting source of dietary and curative compounds. This newly found interest shifted the breeding paradigm from agronomic to health-related purposes. New breeding programs are aiming to develop "specialized" genotypes with high levels of bioactive compounds, vitamins, dietary fibers, and oils.

Keywords

Avena sativa · Oats · Beta-glucans · Avenanthramides · Breeding for HR molecules

1 Introduction

Oat is a cereal grain from the *Poaceae* grass family that originated in the Mediterranean area, also named "brome" and "wild wheat" (Tang et al. 2022). It is an important multipurpose cereal, cultivated for grain, food, feed, fodder, and straw, alone, in mixture, or as a dual-purpose crop. Oats are primarily used as livestock feed that accounts for more than 70% of the total world's usage. Recently, an increased interest in oats arose due to the unique properties of the grain and the nonfood-related applications. In parallel, an increased oat consumption as food has been vastly observed, reaching, in some countries (e.g., the United Kingdom), up to 70% of total oat grain usage. Several products containing oats became of daily use such as breakfast cereals (oatmeal and biscuits), beverages (oat raw or fermented milk), bread, and infant foods.

This cereal plays a key role both in production and sustainable agricultural systems due to its importance in crop rotation and its economic benefits compared to barley production. It is, in fact, a low-input cereal growing relatively well on marginal land or in unsuitable conditions for wheat, barley, or maize production.

Oats are mostly cultivated between 35° and 60° latitude, which correspond to cool and moist climate. Spring-sown oats are the most cultivated in cold environments, while autumn-sown oats (winter oats) are grown mostly in mild climates, such as the United Kingdom, Southern Europe, and the Mediterranean areas. Winter oats have high yields, deep roots, long growth cycle, and early maturity; on the other hand, they lack winter hardiness, making them unfit to be cultivated in very cold areas. In addition, in southern areas, winter sowing allows oats to partially avoid harsh late summer temperatures, even though drought and diseases are still the major causes of yield instability. In spite of this, thanks to new, high-yielding and stress-resistant cultivars, it is gaining popularity also in southern subtropical regions (Sánchez-Martín et al. 2014).

Oat ranks sixth in world cereal production statistics, following wheat, maize, rice, barley, and sorghum (Ahmad et al. 2020).The numbers of total world production in the last years are the following: cultivation area of 9.5 million hectares, yield of 2.4 metric tons per hectare, and production of 23.5 million metric tons. The global oats market amounted to 4.90 billion dollars in 2018 and is estimated to grow by 5.5% over the next 5 years (Kouřimská et al. 2021). During the marketing year 2020/21, approximately 4.72 million metric tons of oats have been yielded in Russia, the world's leading producer, followed by Canada, Spain, Australia, Poland, and China (Fig. 1).

Despite the yield increase of the last 50 years, obtained through new varieties and agronomic practices improvement, the share of minor cereals, that is, oats, rye, and triticale, decreased worldwide. For instance, the total cereal area in Europe downsized from 30% to 18% between 1961 and 2010 (http://faostat.fao.org/), while that of

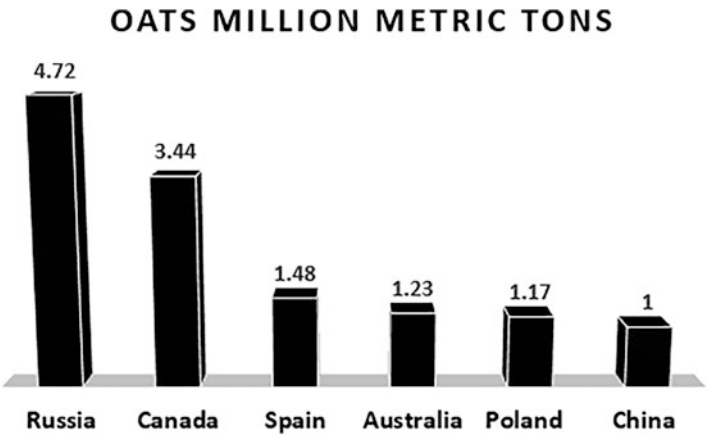


Fig. 1 Leading oats producer countries (Meenu et al. 2021)

the major cereals increased. This proliferation of maize and wheat monocultures means a greater dependence on these crops and consequently higher social and economic vulnerability to risks of soil-borne disease buildup and weed problems. Crop diversification is, therefore, supported, and oats, as low-input crops with higher nutritive efficiency and lower pesticide requirements, are highly appropriate for sustainable cropping systems and of great relevance to agricultural policies development. On the other hand, oats yield is stagnant since the 1980s; therefore, improved oat lines with higher yield and of better quality are urgently required to ensure competitiveness in modern agriculture.

2 Nutritional Composition of Oat

Oat contains valuable and unique macro-, micro-, and phytonutrients. The grains are rich in soluble fibers (mainly β-glucan), polyphenols, galactolipids, and are relatively high in protein content (15–20%) (Kumar et al. 2021), which is quite different when compared with that of other main cereals. It is mainly composed of globulins, prolamins, albumins, and glutelins, which have a superior amino acid profile rich in lysine and threonine. For this reason, oats-based products could be considered ideal plant-derived protein sources for both animals and humans. In addition, oat is an important source of phenolic acids being the only cereal that contains avenanthramides, a group of phenolic alkaloids with beneficial effects on cardiovascular diseases and colon cancer prevention (Dimberg et al., 1993; Soycan et al. 2019). The presence of polyphenols combined with other antioxidant compounds is reported to exert beneficial effects on the human health (Ryan et al. 2011).

Oat carbohydrates content is 99% polysaccharides plus a low fermentable fraction of mono- and oligosaccharides (Bouchard et al. 2022). Starch and fiber (soluble and nonsoluble) are the principal polysaccharides found in oats. Regarding the on-starchy fraction, β-glucan is the most promising water-soluble dietary fiber.

Moreover, oat bran is an excellent source of B complex vitamins (B1, B2, B3, and B6) and tocopherol (vitamin E). The lipidic amount of whole-grain oats is doubled compared with other cereals and mainly consists of unsaturated fatty acids (65%) such as linoleic, oleic, and linolenic acids.

In the light of the above information, we can assume that the peculiar oat grains composition (Table 1), a bran rich in minerals, vitamins, antioxidant, soluble dietary fibers, and an endosperm and aleurone layer rich in proteins, are eligible features to consider oat among the healthiest cereals.

To give a general overview:

- Oats are a good source of proteins, with favorable amino acid contents, high in starch, fats, and vitamins, making them of high nutritional value. Oats have in fact the highest oils (6–12%) and protein (12–20%) contents in de-hulled grains among cereals.
- Oats are rich in beta-glucans: The presence of these polysaccharides resulted in the approval of a health-related claim from the EFSA and the FDA given their

Table 1 Main oat grain nutrients, their variation ranges, and properties, as inferred from recent literature

Nutrients	Total content	Principal compounds	Properties	References
Fiber	10–15.4 g/100 g	Nonsoluble: lignins, cellulose, hemicellulose Soluble: β-glucan (1.73–5.7 g/100 g)	Antidiabetic, antibacterial, cardiovascular protection	Raguindin et al. (2021), Zhang et al. (2021), Chen et al. (2021), Bouchard et al. (2022)
Saponin	49.6–443 mg/kg	Avenacins, avenacosides A (37.7–60.6%) and B (13.8–55.2%)	Antioxidant, lowering cholesterol, affecting immune systems	Shi et al. (2004), Yang et al. (2016), Raguindin et al. (2021)
Phenolics	180–576 mg/100 g	Alkaloid A/p-coumaric acid, alkaloid B/caffeic acids, alkaloid C/ferulic acid (max 149.36 mg/100 g) lignans		Zhang et al. (2021), Raguindin et al. (2021), Soycan et al. (2019)
Flavonoids	n.a.	Quercetin (max 8.9 mg/100 g in husked oats) Rutin (max 0.47 mg/100 g)		Raguindin et al. (2021)
Phytosterols	35–68.2 mg/100 g	β-Sitosterol (59.1–64.9%), campesterol (7.6–9.1%)	Anti-inflammatory, antioxidant, anticancer, preventive and therapeutic for hypertension and coronary heart disease	Raguindin et al. (2021)
Avenanthramides	0.5–71.85 mg/100 g	AVA 2p, AVA 2f, AVA 2c (65–70%)	Antiproliferative, anti-inflammatory, antioxidant, antiatherogenic	Raguindin et al. (2021), Soycan et al. (2019), Liu and Wise (2021)
Fatty acids	6.25–7.5 g/100 g	Unsaturated (65%): linoleic acid, oleic acid, linolenic acids Saturated: myristic acid, palmitic acid, stearic acids	Cardiovascular protection, antiperoxidation of fat	Chen et al. (2021), Bouchard et al. (2022), Kim et al. (2021a, b)
Protein	15–20%	Globulins (50–80%), prolamins (avenins 4–15%), albumins (1–12%), glutelins (10%)		Boukid (2021), Kumar et al. (2021)
Amino acid		Lysine (675 mg /100 g), threonine		Tang et al. (2022)

(continued)

Table 1 (continued)

Nutrients	Total content	Principal compounds	Properties	References
Vitamins		B1/thiamine (0.73 mg/100 g), B2/riboflavin (0.13 mg/100 g), B3/niacin (0.88 mg/100 g), B6 (0.1–0.22 mg/100 g), E/α-tocopherol (0.45–1.2 mg/100 g)	Antioxidant	Chen et al. (2021), Bouchard et al. (2022)
Element and minerals in trace		Calcium (45–58 mg /100 g), phosphorus (325–734 mg/100 g), iron (3.5–5.41 mg /100 g), zinc (3.11–3.64 mg/100 g), manganese (3.63–5.63 mg/100 g), chromium		Chen et al. (2021)

cholesterol-lowering effects. Several other claims, specific to oats and related to reducing the impact of chronic diseases (such as type 2 diabetes, obesity, hypertension, immune-related diseases), are under development. Moreover, oats display higher percentages of soluble beta-glucans in comparison with barley and contain other dietary fibers, notably arabinoxylans.

- Recent investigations on health implications demonstrated that oats' nutritional benefits to human diets go well beyond those currently recognized. Key molecules in this research include polar lipids, antioxidants, and minor bioactive compounds such as avenanthramides.

Because of the grain composition, an increase in oats use for human nutrition has been observed, especially in Europe and North America. The expanding knowledge of the grain composition and its potential end uses are leading to the proliferation of oat-derived products. Table 1 reports the main oat grain nutrients and their ranges of variation depending on cultivation environment, genotypes, and their interaction.

3 Growing Importance in Chronic Diseases and Malnutrition Prevention

Increasing evidence suggests that dietary improvements are important strategies to prevent chronic diseases. Oats possess a variety of health benefits and pharmacological properties. They, potentially, exert antioxidant, anti-inflammatory, antidiabetic, and anticholesterolemic effects. These features could be exploited in functional ingredients production and innovation of traditional foods.

Oat product intake is associated with a reduction in serum cholesterol levels, hence modulation of cardiovascular disease risk (Joyce et al. 2019; Sun et al. 2019). Flavonoids, from whole-grain oat, improve serum lipid profile and decrease lipid deposition contrasting hyperlipidemia (Ren et al. 2021). Studies demonstrated that oat consumption can significantly reduce insulin response, fasting blood glucose levels, and postprandial hyperglycemia incidence.

Beta-glucans' health benefits have been clearly demonstrated. Epidemiological investigations constantly denoted that diets rich in whole-grain products and fibers are associated with a decreased risk of chronic disorders such as type II diabetes, cardiovascular diseases, and cancer.

Furthermore, oats contain plenty of primary and secondary metabolites, ranging from proteins (glutamine in particular) to microelements (copper, iron, selenium, zinc) and polyphenols. These components, together with beta-glucans, exert an immunomodulating role, that is, are able to stimulate the innate and adaptive immune system's response and can inhibit the growth of various bacteria, viruses, fungi, and parasites.

Moreover, oats wield a positive effect on gut's microbiota and related metabolites (Chen et al. 2021) thanks to their dietary structure, resulting in amelioration of chronic problems and improved quality of life. Oat increases *Bifidobacteria* and *Lactobacillus* growth, which exert an antitumoral effect due to short-chain fatty

acids formation, such as butyrate, an important metabolite for intestinal microbial fermentation of carbohydrates (Ren et al. 2021).

Oats consumption can regulate intestinal transit times and increase butyrate production and/or other fecal short-chain fatty acids synthesized by the gut microflora. Patients suffering from inflammatory bowel disease, ulcerative colitis, colorectal adenoma, or cancer can, therefore, benefit from a regular dietary intake of oats.

Of great interest are the different oil fractions and components, for example, polar lipids. Such polar lipids can be used to produce liposomes, which have been demonstrated useful to increase satiety and improve intestinal health (Härröd and Larsson 2011). The studies of Ohlsson et al. (2014) and Hossain et al. (2021) showed that polar lipids regulate hormones involved in human appetite. In the frame of Scan Oats consortium (www.scanoats.se), the combined consumption of lipids and β-glucans has been studied, finding that these two classes of molecules can lower the glucose response, but no synergistic effects have been demonstrated (Cloetends 2022).

Thanks to their fibers and lipids content, oats effectively reduce obesity and are efficient indexes of serum lipid levels and liver function.

Oats and oat products-positive effects are attributed to their fiber amount (especially β-glucan), but also to protein content, balanced amino acids profile, and bioactive substances such as starch, polyphenols, saponins, avenanthramides, and flavonoids. Ingredients derived from oat protein isolates and concentrates could be an added value in foodstuff, especially if considered in combination with the unique functional properties of this cereal (Kumar et al. 2021).

Oats are rich in selenium, involved in DNA repair, and associated with a reduced risk for cancer, especially colon cancer.

Several European countries, the United States, and Canada currently permit oats to be included as an ingredient in gluten-free diets, provided that the gluten contamination is below 20 ppm. Very recently, the proposal to remove oats as priority allergen in Codex Alimentarius has been submitted to FAO/WHO (Tye-Din 2022). Multiple studies reported in fact that oats' prolamin storage proteins or "avenins" do not contain any of the known coeliac disease epitopes from gluten of wheat, barley, and rye (Gilissen et al. 2016; Smulders et al. 2018). Others found that oats have a very low or null immunogenicity, depending on the different cultivar (Comino et al. 2015; Kosová et al. 2020). Recent genomic studies suggested that *A. sativa* has low copy number of genes encoding celiac disease epitopes, low occurrence of highly immunogenic proteins, and low proportion of avenins within total oat proteins (Tye-Din 2022; Kamal et al. 2022). Long-term studies confirmed the safety of oats for celiac disease patients and the positive health effects of oat products in a gluten-free diet. Controlled avenin feeding studies concluded that protracted oat ingestion should not be harmful, even if a small sensitive subset of patients are likely to exist (Tye-Din 2022). Inclusion of oats in a gluten-free diet might be valuable due to their nutritional and health benefits, provided that no contamination from other small grain cereals can be guaranteed in the supply chain (Comino et al. 2015). Wheat, barley, or rye contamination is in fact a major problem in oats' conventional production chains.

Oat is different from other cereal grains as the whole de-hulled kernel is used. Many food products are available, for example, porridge or oatmeal, hot cereals, bread, biscuits, infant food, muesli, or granola bars and dairy substitutes, such as oat milk, yogurt, and ice cream. The technological processes applied to the production of different oat-based food can impact nutritional properties. For example, the three thermal treatments coupled to a starvation step used for the preparation of traditionally oat-based Chinese food impact the β-glucan and protein MW and on GI (Hu 2022). Starting from the numerous positive effects on human health, the consumption of oats and derived products can be encouraged. The diffusion of these products should be based on informative campaigns toward consumers, backed by robust scientific results. This will provide new markets for higher added value products and higher profitability in the oat-related agrofood sector.

4 Health-Related Molecules Unique to Oat

Oats gained a great deal of interest as a healthy and highly nutritive food, rich in secondary metabolites of great functional value. Two classes of HR compounds' characteristics of oats are beta-glucans and avenanthramides (AVNs). β-Glucans have hypocholesterolemic, hypoglycemic, antitumor, immunomodulatory, antioxidant, and anti-inflammatory activities (Kim et al. 2021a, b). AVNs have strong and widely demonstrated anti-inflammatory activities. Tranilast, a commercially available antiallergic drug similar to AVN, has been demonstrated to potentially prevent exacerbation of COVID-19 by affecting different pathways such as NLRP3 inflammasome, signaling pathways, cytokines and chemokines, and cell adhesion molecules (Saeedi-Boroujeni et al. 2021).

4.1 β-D-Glucans

(1 → 3), (1 → 4) β-D-glucans (Fig. 2) are nonstarchy polysaccharides found in aleurone, subaleurone, and starchy endosperm cell walls. They represent the major component of the soluble fiber in oats. They are composed of glucose monomers in long linear glucose polymers, linked by β (1 → 4) (70%) and β (1 → 3) (30%) glycosidic bonds (Bouchard et al. 2022).

The difference between soluble and insoluble β-D-glucans is based on the ratio of the two bonds. In fact, the presence of β (1 → 3) gives greater flexibility to the molecule; consequently, more solubility and viscosity. Bonds formation is influenced by genotype and growing environment (Redaelli et al. 2013).

Moreover, about 90% of glucose units are organized in trimers and tetramers joined by β (1 → 3) bonds; their ratio influences structural variability and consequently the differences in some physical properties. The ratio between cellotriose and cellotetraose is characteristic of every different cereal with a smaller value for oats (1.9–2.4), a higher percentage of cellotetraose makes the molecule less soluble (Bouchard et al. 2022; Bai et al. 2019; Wang and Ellis 2014). Higher molecular

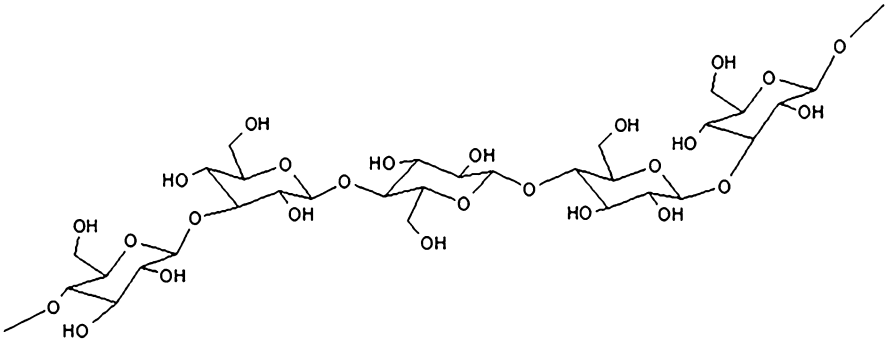


Fig. 2 Chemical structure of (1 → 3), (1 → 4) β-D-glucans

weight and lower cellotriose/cellotetraose ratio β-glucans tend to have higher viscosities than those with lower molecular weight and higher cellotriose/cellotetraose ratio (Bouchard et al. 2022).

The biosynthesis of (1 → 3), (1 → 4) β-D-glucans depends on an enzymatic complex consisting of a cellulose-synthase-like enzyme (*Csl* gene), which leads to the formation of cellobiose, units of cellodextrin, and a glycosyl-transferase enzyme, which adds another glycosyl residue to form cellotriose and higher odd-numbered chains units (Izydorczyk and Dexter 2008). *CslF6* is the major gene responsible for the biosynthesis of (1,3; 1,4) -β-d-glucans. It is highly expressed in different tissues, especially in the developing endosperm, and produced in the Golgi apparatus, then is channeled through the secretory pathway to the plasma membrane. Changes in the *CslF6* gene, of even a single amino acid, can alter the fine structure of the (1,3;1,4)-β-D-glucans (Chang et al. 2021).

Oat is rich in both soluble, particularly β-glucan, and insoluble fiber, particularly cellulose, lignin, and hemicellulose. A diet high in fiber is known to be beneficial in the prevention of many diseases.

4.1.1 Reduction of Cholesterol and Postprandial Glucose in the Blood

Soluble β-glucan can increase the viscosity of the food bolus by delaying stomach emptying, improving intestinal filling, and slower absorption of nutrients. More voluminous stools accelerate intestinal motility by reducing the time of exposure of the intestinal wall to irritants and carcinogens present (Singh et al. 2013).

The increase in gastrointestinal viscosity also causes cholesterol and bile acids to be trapped, reducing their absorption; consequently, the bile acids necessary for digestion must be synthesized from cholesterol by reducing their concentration. β-Glucans have been shown capable of reducing the total and LDL-cholesterol level, in the blood of normo- or hypercholesterolemic subjects, reducing the risk of heart disease. The physiological response depends on solubility, concentration, molar mass of β-glucan, and molecular weight, all contributing to increase viscosity (Singh et al. 2013; Zhang et al. 2021).

β-Glucan solutions' high viscosity is able to reduce postprandial glucose by interfering with the transfer of released glucose to enterocytes. Another pathway is oat β-glucans consumption by the intestinal microbiota in the colon; hence, they are not digested in the upper gastric tract. Fermentation produces short-chain fatty acids, such as propionic, butyric, and acetic acid, which can regulate the expression of insulin-sensitive glucose transporter type 4 (GLUT-4) imputed to maintaining glucose concentration gradient across cell membranes and allowing the transport of glucose (Zhang et al. 2021).

The reduction of postprandial glucose and therefore of inflammation caused by glucose is consequently reducing the general inflammatory state (Chen et al. 2021).

The hypoglycemic function of β-glucans is influenced by its molecular weight, a higher value is associated with a greater decrease in fasting blood glucose levels and better control of blood glucose levels after meals. β-Glucan can lower the glycemic index (GI) of foods by delaying digestion and absorption of starch as a result of amylase activity reduction, thus obtaining a hypoglycemic effect, and the GI value is decreased with increasing molecular weight and viscosity of β-glucan (Tang et al. 2022).

Scientific evidence about these beneficial effects of oats and its β-glucans led to the authorization of specific health claims that can be reported on labels of foodstuff constituted of oats and barley in order to clearly communicate benefits and limitations associated with consumption to the consumer.

The European Food Safety Authority (EFSA) authorized the claim that β-glucan contributes to the maintenance of normal cholesterol levels in the blood and the beneficial effect is obtained with a daily intake of 3 g of beta-glucans from oat or oat bran, barley, or barley bran or mixtures of these (EFSA 2009; Commission Regulation EU 2012). In a subsequent opinion, EFSA established that a reduction in cholesterol can diminish the risk of (coronary) heart disease and the amount of β-glucan to be taken throughout the day must be part of a balanced diet (EFSA 2010).

The use of the claim has also been approved by the Food and Drug Administration (FDA) in the United States, indicating that soluble fiber, 3 g or more per day of β-glucan from either whole oats or barley, or a combination of whole oats and barley, can reduce the risk of coronary heart disease through the intermediate link of blood cholesterol or total LDL cholesterol in the blood (Code of Federal Regulations, 21CFR101.81), recognized also by Health Canada and Food Standards Australia New Zealand.

Another claim approved by EFSA and based on scientific evidence is that the consumption of β-glucan from oats or barley as part of a meal contributes to the reduction of high blood glucose after a meal. The claim can be used only for food containing at least 4 g of β-glucan from oats or barley for every 30 g of carbohydrates available in a quantified portion as part of the meal (EFSA 2011).

4.1.2 Effect on the Immune System, Cancer Prevention, and Antimicrobial Activity

Oats β-glucans improve immunity and anticancer activity, and moreover, can kill the malignant cancer cells, sarcoma cells, and melanocytes. The inhibition exerted on

different types of intestinal, liver, and breast carcinomas is similar to anticancer drugs but without any side effects (Tang et al. 2022).

β-Glucans have an immunostimulant effect altering the microflora of the colon with the production of short-chain fatty acids (Singh et al. 2013).

The fermentation of soluble fiber with the production of short-chain fatty acids, in particular butyric acid, favors the development and colonization of probiotics. Short-chain fatty acids improve cell proliferation of the colon mucosa, reducing the risk of cancer (Singh et al. 2013).

Short-chain fatty acids (SCFA), bacterial metabolites of dietary fibers, stimulate the production of mucus and antimicrobial peptides, increase the expression of tight junction proteins, and modulate immune system (Chen et al. 2021). Multiprotein junctional complex plays a key role in osmose balance maintenance and transcellular transport of specific molecules.

Fermentable fibers can reduce the amount of mucin used during fermentation; mucin acts as a barrier against pathogenic microorganisms, and fiber contributes to intestinal microbiota maintenance by reducing the chance of opportunistic pathogens to prevail (Chen et al. 2021).

β-Glucan concentration has a positively proportional antibacterial effect. Low molecular weight β-glucan entering microbial cells would cause their lysis and death (Tang et al. 2022).

4.1.3 Blood Pressure Reduction

β-Glucans can reduce blood pressure, likely due to their viscosity. Insulin resistance, an important indicator of hypertension, has been linked to β-glucans viscosity. The intake of viscous fibers influences renal sodium absorption and transmembrane ion transport, causing a reduction in blood pressure (Bai et al. 2019).

4.1.4 Antioxidant and Anti-inflammatory Activity

Low molecular weight β-glucans transfer hydrogen from molecules that can act as free radical quenchers under physiological conditions (Bai et al. 2019). They have free radical "scavenger" activity and the ability to relieve inflammatory conditions.

4.2 Avenanthramides

Oats are rich in phenolic compounds that are interesting for their high antioxidant capacity and the potential health benefits (Boz 2015). Phenolic compounds consist of aromatic rings with one or more hydroxyl groups (Bouchard et al. 2022). Oats contain mainly simple phenols such as phenolic acids, flavonoids, and avenanthramides compounds (Tang et al. 2022). Avenanthramides were originally identified as phytoalexins produced by the plant when exposed to pathogens (Perrelli et al. 2018).

The avenanthramides (AVAs) are a group of phenolic alkaloids typical of oats (Fig. 3); they are low molecular weight-soluble phenolic compounds (Boz 2015)

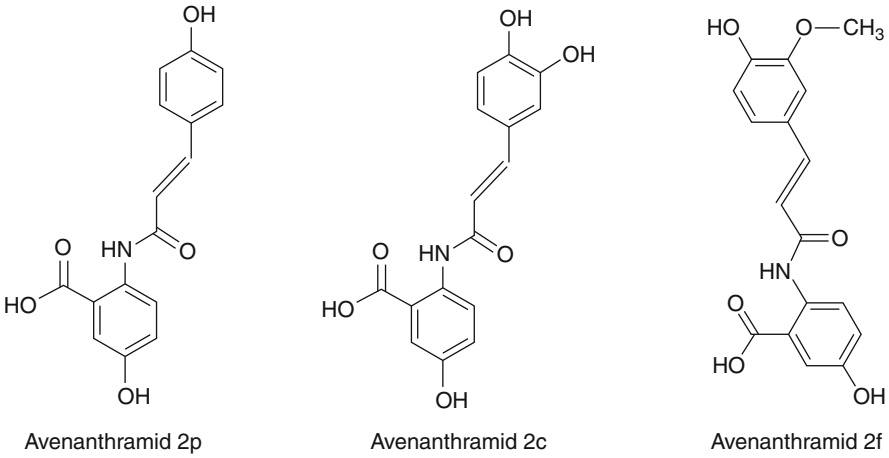


Fig. 3 Chemical structure of avenanthramides

consisting of anthranilic acid and hydroanthranilic acid linked to hydroxycinnamic acids by amide bonds (Singh et al. 2013).

Oats contain approximately 40 different types of AVAs, the predominant are esters of 5-hydroxyanthranilic acid with p-coumaric (AVA-A or 2p), ferulic (AVA-B or 2f), and caffeic (AVA-C or 2c) (Boz 2015). Among cereals, only oats contain AVAs; they can be found in almost all fractions of the kernel but particularly in bran (Bouchard et al. 2022) and aleurone layers (Tang et al. 2022).

While the mechanisms of biosynthesis of the main avenanthramides in oats AVA-A, AVA-B, and AVA-C are not yet fully understood, studies identified three different types of genes that encode 4-coumarate-CoA ligase (4CL), hydroxycinnamoyl-CoA, hydroxyanthranilate N-hydroxycinnamoyl transferase (HHT), and caffeoyl-CoA O-methyltransferase (CCoAOMT) enzymes, all involved in the biosynthesis process of avenanthramides. 4CL oats appear to convert p-coumaric, caffeic, and ferulic acids in their CoA thioesters, while oats' HHT are responsible for biosynthesis of AVA-A and AVA-C by the condensation process of hydroxyanthranilic acid, acyl acceptor, with p-cumaroyl-CoA and caffeoyl-CoA, acyl donor. AVA-B is synthesized by methylation of the hydroxyl group at position 3 of the aroyl group in AVA-C by the CCoAOMT enzyme (Li et al. 2019).

4.2.1 Antioxidant, Anti-inflammatory, and Antiatherogenic Activity

Avenanthramides showed antioxidant potential significantly higher than other simple phenols. Antioxidants protect cells from the oxidative damage and help prevent several chronic diseases caused by reactive oxygen species (ROS) generation. Antioxidant activity was found to be in AVA-C> AVA-B> AVA-A. AVAs inhibit calcium-induced low-density lipoprotein (LDL) oxidation in a dose-dependent manner and are shown to act synergistically with vitamin C and other antioxidant compounds. During physical activity, taking avenanthramides decreases ROS

production and related lipid peroxidation. The antioxidant potential of AVAs could be related to an effect on antioxidant enzymes, such as superoxide dismutase (SOD) and glutathione (GSH) peroxidase (Tripathi et al. 2018). AVAs are bioavailable, and their antioxidant effect on LDL reduces the risk of cardiovascular diseases (Singh et al. 2013).

AVAs also inhibit inflammatory processes that play a significant role in many diseases. AVAs' anti-inflammatory activity was detected in endothelial human aortic cells with an evident reduction of various types of molecules involved in the attachment of monocytes (blood immune cells) to the arterial walls, responsible for inflammation and stiffening, first step in the development of atherosclerosis (Singh et al. 2013). AVAs reduce production of pro-inflammatory cytokines (interleukin IL-6, IL-8, and monocyte chemoattractant protein (MCP-1)) that carry the immune cells at the stimulation site and inhibit vascular endothelial cell expression of adhesion molecules, including ICAM-1 (intracellular adhesion molecule-1), VCAM-1 (vascular adhesion molecule-1), and E-selectin (Tripathi et al. 2018; Singh et al. 2013).

Colloidal oat flour has been found to be effective as an anti-inflammatory and anti-itch in dermatitis treatment, with avenanthramides responsible for these beneficial effects. The anti-inflammatory effects of AVAs have been demonstrated in human keratinocytes. AVAs reduce the IL-8 production, and the anti-inflammatory action occurs through the NF-κB pathway. NF-κB is one of the very crucial regulatory transcription factors (Tripathi et al. 2018).

4.2.2 Antiproliferative Activity and Postprandial Glycemic Response Control

AVAs, including AVA-C and its methylated derivative, can inhibit the proliferation of colon, breast, and smooth muscle vascular cancer cells. They cause cell cycle blockage in the G1 phase by upregulating the p53-p21cip1 pathway and inhibiting the retinoblastoma protein (pRB) phosphorylation with the possible activation of an apoptosis process (Perrelli et al. 2018).

Glucose uptake can be mediated in humans by intestinal solute carriers SGLT1 and GLUT2. In a proof-of-concept study, Joseph (2022) showed that AVAs can inhibit glucose uptake in a dose-dependent manner in CaCo2 cell system.

5 Genetic Resources of Health-Related (HR) Genes

The genus *Avena* includes up to 30 recognized species, with different ploidy levels that includes diploids ($2n = 14$), tetraploids ($2n = 28$), and hexaploids ($2n = 42$). As reviewed by Loskutov (2005, 2008), the genus comprises the two sections of *Aristulatae* and *Avenae*, including both wild and cultivated species (Table 2).

Among cultivated, *A. strigosa* is a diploid with $2n = 14$ (genome As), whereas *A. abissinica* is a tetraploid that carries the two A and B genomes ($2n = 24$).The most cultivated *A. sativa* and marginally cultivated *A. abyssinica* are hexaploids with 42 chromosomes, presenting three different sets of nuclear genomes A, C, and D.

Table 2 Main oat species according to Loskutov (2005)

Section	Wild species	Cultivated species
Aristulatae	*A. claudia*	*A. strigosa*
	A. prostrata	*A. abyssinica*
	A. damascena	
	A. longiglumis	
	A. wiestii	
	A. hirtula	
	A. barbata	
	A. vaviloviana	
Avenae	*A. fatua*	*A. byzantina*
	A. occidentalis	*A. sativa*
	A. ventricosa	
	A. bruhnsualis	
	A. canariensis	
	A. magna	
	A. murphyi	
	A. canariensis	
	A. insularis	
	A. sterilis	
	A. ludoviciana	

Three gene pools can be recognized inside the genus, classified as primary, secondary, and tertiary depending on the interfertility with cultivated hexaploid oat, as reviewed by Ociepa (2019). The primary gene pool (GP-1) collects accessions, including landraces, breeding lines, modern and obsolete cultivars, belonging to the same species as the cultivated oat. The secondary gene pool (GP-2) includes species that are still crossable with cultivated oats, even if a limited number of fertile hybrids are produced. The tertiary gene pool (GP-3) includes relatives that are so distant that the crossing would require very specialized techniques such as embryo rescue, bridge crosses, and induced polyploidy.

The primary gene pool is wide and collects wild and cultivated hexaploid species, interfertile and without any recombination restrictions. New genes from wild species have been introduced in cultivated forms mainly to improve agronomic traits. *A. sterilis* has been an interesting donor of resistance genes against fungal diseases such as powdery mildew, crown rust, and stem rust. *A. sterilis* contributes even to the improvement of winter hardiness, grain yield, drought tolerance, and green mass. This same species, *A. sterilis*, is a potential donor of desirable quality-related traits, such as large grains, high protein content, balanced amino acid composition, and high oil and beta-glucan contents. *A. fatua* has also been included in oat-breeding programs as genes donor for reduced plant height, abiotic stress resistance, stem and crown rust resistance, smuts, and viroses. *A. fatua* and *A. ludoviciana* are useful sources of resistance to crown rust. Aside from agronomic traits, these two species are considered donors of important genes for health-related characters, such as high protein and oil content.

Exploitation of *A. sterilis* in breeding programs gave cultivated varieties, for example, Starter (characterized by high-protein content), Ozark (cold resistant), and Jay (resistant to crown rust).

The secondary gene pool includes the tetraploids *A. murphyi, A. maroccana,* and *A. insularis*.

The crosses between *A. sativa* and these species give partially fertile hybrids, whose fertility level can be increased by backcrossing. *A. murphyi* could be a potential source of resistance to powdery mildew and crown rust, but also donor of high protein content and high groat oil content. *A. magna* can contribute to both quality-related traits (grain size, high contents of protein, lysine, and oil) and biotic stress resistance (powdery mildew and crown rust). Crosses with *A. maroccana* resulted in varieties such as Amagalon, resistant to crown rust, and CDC Bell for green feed or oat hay production.

In the tertiary gene pool are included all diploid and tetraploids species. Resistance to crown and stem rust, Septoria, powdery mildew, together with high oleic acid and fiber contents, are interesting features of several diploid and tetraploid wild species. However, cross-incompatibility complicates the gene transfer from diploids and tetraploids to hexaploids. This problem can be mitigated using backcrosses, mutants, and genetic intermediates. The varieties Hinoat, Gemini, and Foothill, all resistant to crown rust, have been obtained by crossing *A. sativa* with the diploid, marginally cultivated, *A. strigosa*.

Considerable genetic resources exist with large collections of landraces, wild relatives, and cultivars. According to De Carvalho et al. (2013), currently, 116 major GeneBanks conserve oats' genetic resources. The major collections of oat landraces are located in Canada (Plant Gene Resources of Canada (PGRC)) at the Saskatoon Research and Development Centre (Saskatoon, Saskatchewan), Russian Federation (Federal Research Center N. I. Vavilov All-Russian Institute of Plant Genetic Resources, VIR), the United States (NSGC), and Germany (Leibniz Institute of Plant Genetics and Crop Plant Research, IPK, Gatersleben). The collections include adapted cultivars, old landraces, marginally cultivated oats such as *Avena strigosa*, *Avena abyssinica*, *Avena brevis,* and *Avena nuda*, crop wild relatives, and genetic stocks bearing defined traits. Different Gene Bank information systems have been developed (e.g., GRIN-CA/GRIN-Global-CA, GRIN-USA, EURISCO) that allow national and international clients and the public to inspect and access to the GeneBank holdings.

Significant sources of genetic variation for oat breeding are therefore mainly available in European and American GeneBanks and breeding programs, coherently with the spreading areas of this crop.

Cultivated *A. sativa* was most likely first domesticated in Europe ca. 2000–3000 BC (Zohary et al. 2012) and then spread from the East to Central–North Europe in the late Bronze Age. In the western part of the Mediterranean region originated A. *byzantine,* the main cultivated form in North Africa and Spain.

In the sixteenth century, the two *A. sativa* and *A. byzantina* species were introduced to North America. *A. sativa*, more suited to spring sowing, was mainly spread in the northern regions, whereas *A. byzantina* was used for autumn sowing in the Southern United States. In the twentieth century, new breeding programs

built on the crosses between the two species and gave many varieties currently cultivated in these environments.

6 Oat Breeding for Quality and Health-Related Traits

In the review of Loskutov and Khlestkina (2021), it is clearly reported how the oats' breeding paradigm has been recently switched from agronomic objectives to health-related ones. The traditional breeding has been directed to improve yield and agronomical traits, such as abiotic and biotic stress resistance, lodging resistance, and growth habit, together with a few quality-related traits, namely protein and starch content. During the last years, oats turned out to be a very interesting source of dietary and curative compounds. New breeding programs are aiming to develop "specialized" genotypes with high levels of bioactive compounds, vitamins, dietary fibers, oils, etc. This is in line with the Rome Declaration on Nutrition, in which integrated strategies are indicated to eradicate all forms of malnutrition, that is, not only undernutrition, but even micronutrient deficiencies, and obesity. One of the strategies is optimizing the nutritional value of crop and derived raw materials. Starting from the prominent role of small grain cereals in human nutrition, the importance of new biofortified cereal varieties, improved in essential amino acids, fatty acids, vitamins, minerals, etc., is undeniable (Shelenga et al. 2021).

Traditional breeding approach to biofortification involves crosses between genotypes high in target compounds and varieties with superior adaptability and agronomic performances (Fig. 4). This approach requires to find interesting genotypes by

Fig. 4 Scoring oat plants in traditional breeding program

evaluating the existing variability not only in accessions belonging to primary gene pool, but even in wild relatives. The breeding can then be accelerated combining traditional techniques with marker-assisted selection, next-generation breeding, and improved phenotyping strategies for specialized compound identification and quantification.

In traditional breeding programs, the evaluation of phenotypes is classically done in several environments; selection and recombination are based on the resulting data plus pedigree information, when available. Conventional crossbreeding in oats has been, up to now, based on relatively narrow genetic diversity (He and Bjørnstad 2012; Tinker et al. 2009). However, opportunities to significantly enhance genetic gain in crop breeding by combining phenotypic selection with precise molecular breeding approaches are rising. The other pillar is the exploitation of wider germplasm available in GeneBanks: genetic resources belonging to the different gene pools host interesting traits for crop improvement in various farming systems and under changing environmental scenarios. Table 3 summarizes the main targets for oat improvement.

The main task of traditional breeding has been the yield increase. However, at the beginning of the twenty-first century the yield increase obtained with the new oat varieties was just over a third of the increases recorded in other small grain cereals (Menon et al. 2016). Moreover, even smaller breeding efforts have been invested in naked oats, despite their potential both from economic and nutritional point of view. There is therefore a large genetic gain margin to improve not only the yield, but also the nutritional characteristics to utilize oat as an ingredient in functional food.

The improvement of such polygenic traits can be supported by molecular breeding strategies, such as molecular-assisted selection (MAS) and mutagenesis (TILLING), supported by QTL mapping and genome-wide association studies (GWAS).

The identification of superior genotypes can be speeded up using molecular markers linked to the trait of interest, but to unravel the genetic basis of complex traits it is necessary to associate genotypic information with the corresponding phenotypic data. Association strategies are based on three main pillars: the phenotyping, the genotyping strategy, and the genetic population used (Isidro-Sánchez et al. 2020b).

The first linkage-based QTL map was developed by O'Donoughue et al. (1995), using 71 recombinant inbred lines from a cross between *Avena byzantina* cv. Kanota and *A. sativa* cv. Ogle and restriction fragment length polymorphisms (RFLP) as molecular marker. Several other maps, at increasing density, have been developed since then. Different mapping populations were developed in combination with several classes of molecular markers such as RAPDs, ISSRs, IRAPs, SCARs, DArTs, and SNPs (Gorash et al. 2017).

The first doubled haploid linkage map for cultivated oats was created in 2008 (Tanhuanpää et al. 2008), whereas Oliver et al. (2013) developed and published the first physically anchored hexaploid oat linkage map. Tinker et al. (2014) developed the first SNP genotyping array for hexaploid oat and a linkage map of naked oats was constructed by Song et al. (2015).

Table 3 Main objectives of oat breeding. Some of the target traits have been already considered by traditional breeding, some others – mainly related to quality – are new ones

Objective	Target trait(s)
Competitiveness in cereal farming systems and supply chains	Yield potential and stability Various end user demands: food, feed, other derived products Plant architecture: dwarfing types to avoid lodging
Host plant resistance to pathogens	Crown rust *Fusarium* head blight (and minimizing mycotoxins in grains) Oat mosaic virus Powdery mildew
Adaptation to stressful environments	Cold tolerance Drought tolerance Deep root systems Competing ability against weeds
Grain quality	High kernel content Ease of dehulling Low proportion of screenings High specific weight Minimum grain blackening High protein and high essential amino acids High oil and high essential fatty acids High total and soluble fiber: β-glucan and arabinoxylan High bioactive phytochemicals: phenolics and terpenoids
Input-use efficiency and mitigating climate change	Nitrogen use efficiency Water use efficiency
Healthy food	Increased β-glucan content Increased antioxidants
Livestock feed	High metabolizable energy by increasing oil content Low lignin husk High yield, protein and oil in naked oats (for ruminants) Low trichome density in naked oats (for facilitating harvest and grain handling)

The first linkage maps have been built on bi-parental populations; however, such materials can have some limits in genetic and phenotypic variation. Other kinds of populations, showing wider variation, are under development, including multiparent advanced generation intercross (MAGIC) and nested association mapping (NAM) populations. These kinds of populations can avoid biasing the selection of marker loci toward those that are polymorphic only between specific mapping parents. Moreover, an increased mapping resolution can be reached thanks to the numerous generations of intercrossing of the original parent plants and their ability to capture a wide range of genetic and phenotypic diversity in a breeding population. Such populations form the perfect basis for quantitative trait locus mapping and MAS in multiple genetic backgrounds.

This strategy has been used to map QTL involved in β-glucan concentration (Newell et al. 2011; Asoro et al. 2013), disease resistance (Gnanesh et al. 2015; Montilla-Bascón et al. 2015), spikelet number (Pellizzaro et al. 2016), heading date (Klos et al. 2016), and frost and lodging tolerance and heading date (Tumino et al. 2016).

The marker development in oat evolved from low-throughput to high-throughput markers. An international consortium of oat researchers partnered with Diversity Arrays Technology P/L (Canberra, Australia) to develop a high-throughput DArT marker platform for oat (Tinker et al. 2009). This collaborative effort successfully developed >2700 DArT markers, of which 1295 were unique and polymorphic in a diverse set of global oat germplasm. This work has more than doubled the number of available oats markers and provided the first high-throughput marker platform in oat; it can be applied in parallel and is based on a rapid and cost-effective commercial service.

A further evolution was reached with single-nucleotide polymorphisms (SNPs) markers, as they are suitable for automation, easy to exchange among laboratories, and to score. Next-generation sequencing (NGS) approach has been addressed in a collaboration of North American and European oat researchers, resulting in the first SNP oat genotyping platform – an Illumina® 6K oat chip containing highly informative SNP markers (Tinker et al. 2014) – which is now available to oat breeders and is the first physically anchored integrated map of the complex oat genome (Oliver et al. 2013) for further use in genetic analyses. This work has reversed the declining trends in oat research by infusing the research community with the genetic resources necessary to develop new and innovative cultivars.

The SNP array has been used to evaluate the genetic diversity and population structures of American and European germplasm, whose genetic base is different (Tinker et al. 2009), and also for linkage and association mapping, thereby advancing genetic analysis and breeding. A problem in identifying informative SNPs in allohexaploid oat can be the occurrence of homeologs from the A, C, and D genomes. A number of bioinformatics tools have been developed to successfully detect and remove sequences with redundant and ambiguous polymorphism allowing efficient discovery and application of SNP markers in oats (Barker and Edwards 2009; Oliver et al. 2011). Moreover, Tumino et al. (2016) proposed to utilize SNP hybridization intensity ratios as continuous variables to better represent the allele frequencies at accession level (bulk). Using intensity ratios, SNP signals were not interpreted as discrete genotype classes but as continuous. The bulk allele frequencies can be statistically associated with phenotypes measured in bulks in GWAS. The same approach has been adopted by Montilla-Bascòn et al. (2015) in a GWAS for crown rust and powdery mildew. The main reasons supporting this strategy are the reduction of genotyping costs and the lowered risk of useful information loss given that SNP probes in allopolyploid species may or may not be subgenome-specific and therefore the genotype calling can be a challenging, time-consuming, and error-prone task.

GWAS exploits natural large genetic resources collections for the identification of genomic regions that are in linkage disequilibrium (LD) with the QTL influencing

traits of interest. This method takes advantage of ancestral recombination events and can provide higher resolution compared to traditional QTL mapping using biparental populations. Association genetics uses, therefore, large experimental populations with densely mapped genotype data and phenotypic data from multilocational field trials. An example of such populations is the Collaborative Oat Research Enterprise (CORE) panel, which consists of 635 single-panicle-derived lines representative of elite germplasm deemed important by 16 active oat breeding programs in Australia, Canada, the United Kingdom, and the United States. A further example is the AVEQ panel, developed to evaluate oat genetic resources for proteins, oils, minerals, β-glucans, antioxidants, and phenolic compounds, as well as for resistance to *Fusarium* infections and cold tolerance, assembling 600 accessions and cultivars from 25 gene banks and 31 breeding programs from 14 European countries.

The analysis is based on the nonindependence of alleles in a population called linkage disequilibrium (LD), defined as the nonrandom association of alleles at two loci. Association between marker alleles and causal alleles arises not from experimental crossing but from historical drift and mutation events. In such studies, associations between genotype and phenotype depend on historical LD broken down by many generations of recombination. For this reason, in GWAS a larger number of markers are required to assure LD between markers and causative alleles throughout the genome, thus enabling finescale mapping. This allows us to map causal loci more accurately than with traditional linkage analysis.

The genetic distance over which LD is maintained in a population determines the resolution of mapping that is possible, and the marker density required for association analyses. A key point is the population structure deriving from admixture, mating system, genetic drift, artificial or natural selection during evolution, domestication, and breeding. Population structure can be responsible for false associations between polymorphic markers and phenotypic trait variations. Several statistical approaches are therefore available to describe population structures and genetic relatedness of the lines in the association panel. A lot of studies showed that population structure in oats is weak (Montilla-Bascón et al. 2015; Newell et al. 2012; Tumino et al. 2017; Winkler et al. 2016) in comparison with other cereals (Hamblin et al. 2010) probably due to admixtures and spring and winter types interbred species (Newell et al. 2012). Consequently, numerous GWAS for oats have been successfully concluded (Asoro et al. 2013; Klos et al., 2016; Foresman et al. 2016; Montilla-Bascón et al. 2015; Newell et al. 2012; Tumino et al. 2016, 2017; Winkler et al. 2016).

NGS combined with the reduction in genome complexity enables genotyping-by-sequencing (GBS) approaches (Bekele et al. 2020). This combines marker discovery and genotyping to produce high-density markers at a relatively low sample cost (Poland and Rife 2012). The development of such tools would enable genomic selection (GS) to be used in oats (Mellers et al. 2020). This has recently been proposed to overcome the limitations of current MAS approaches for complex traits governed by many genes. Instead of analyzing markers individually for their association with a phenotype, GS analyzes all the markers jointly to explain the total genetic variance by summation of all marker effects. These are then used to predict

the breeding value of individuals. The development of high-density SNP maps and of high-throughput genotyping platforms makes GS a practical proposition for oat breeding (Asoro et al. 2011), as already established in other cereals (Poland and Rife 2012).

A very high number of different molecular maps both for hexaploidy and diploid oats have been developed (Blake et al. 2019), and several of these can be found in GrainGenes database (https://wheat.pw.usda.gov/GG3/). The same database hosts reference genomes of hexaploidy oat and of *A. insularis*, *A. longiglumis*, *A. atlantica,* and *A. eriantha*.

Chip-based SNP assays and GBS approaches are ideal genotyping platforms for GWAS and GS, but breeders often require screening large numbers of plants with small numbers of mapped markers associated with key traits of interest. KASPar SNP genotyping answer to this need as it is a flexible and cost-effective system able to discriminate SNPs even in hexaploidy species. KASPar assays can also identify heterozygous individuals, and this can be important in breeding and backcrossing programs, where marker-assisted selection often requires the resolution of heterozygous from homozygous classes.

To map populations and germplasm collections, TILLING (targeting-induced local lesions in genomes) populations, obtained with an advanced mutation breeding method, can be employed. This nongenetically modified technology combines the traditional mutagenesis with high-resolution mutation screening and allows to expand the genetic variability and detect point mutations in individual plants (Parry et al. 2009). In oats, TILLING was applied by Chawade et al. (2010) to obtain mutant seed lines that have been used to study key genes in the lignin and β-glucan biosynthetic pathways. Vivekanand et al. (2014) studied the lignin-level variability inside this same population and in Belinda variety, from which population was obtained, with the aim to identify more digestible mutants.

Finally, González-Barrios et al. (2021) demonstrated how speed breeding and early panicle harvest can accelerate oat breeding cycles. By modulating growth temperatures and photoperiod, a significant acceleration in flowering time has been obtained and acceptable germination levels are present 21 days after flowering. Breeding programs where single-seed descent are used can be therefore greatly accelerated.

6.1 Genetics of FHB Resistance

Mycotoxins are secondary metabolites produced by different types of fungi such as *Aspergillus, Penicillium, Alternaria,* and *Fusarium*. These toxins can enter the food chain through contaminated crops used in food and feed production. The mycotoxigenic molds can invade both agricultural crops in the field and agricultural raw materials during storage and processing (Mielniczuk and Skwaryło-Bednarz 2020). Oat contamination by mycotoxin is of great concern for human and animal health (Perrone et al. 2020). Several mycotoxins can be detected in oats, but the ones considered most important are deoxynivalenol (DON, vomitoxin), zearalenone

(ZON or ZEN), HT-2 toxin (HT2), and T-2 toxin (T2). DON and ZON are produced by several *Fusarium* species, the dominant being *F. graminearum*. HT2 and T2 are produced by several other *Fusarium* species, the dominant being *F. langsethiae* (Edwards et al. 2012). The *Fusarium* species invade the plant at flowering stage causing Fusarium Head Blight (FHB). Legislative limits for the mycotoxins DON and ZON for unprocessed cereals, intermediate products (e.g., flour), and finished products for human consumption have been fixed in several countries.

Screening for sources of resistance to FHB is of key importance for the development of new oat cultivars. However, resistance to FHB is a quantitative trait (He et al. 2013; Bjørnstad et al. 2017), controlled by genes that are still largely unknown. Resistance can be divided into several different components, as reviewed by Hautsalo et al. (2018), resistance against initial infection and spreading of infection (respectively known as type I and II), mycotoxin (type III), kernel infection (type IV), and tolerance (type V). In addition, disease escape mechanisms exist, and are based on morphological features, such as plant height, degree of anther extrusion (Tekle et al. 2020), or physiological characteristics (flowering time). Moreover, mycotoxin accumulation is strictly linked to the fungal species, and infection success is strictly linked to agronomic practices and weather conditions (Hautsalo et al. 2020).

Oat genetic resources have been screened to identify the potential source of resistance (Bjørnstad et al. 2017; Loskutov et al. 2017; Hautsalo et al. 2020; Gavrilova et al. 2021). Screening procedure itself has been demonstrated to be a nontrivial point to obtain reliable indication about resistance (Tekle et al. 2018). A ranking of oat cultivars with respect to the content of DON can be informative when repeated several years in a specific environment, as reported by Tekle et al. (2018) and Yan et al. (2010).

Haikka et al. (2020) evaluated North European germplasms and breeding lines for DON and agronomic traits and compared two strategies, GWAS and genomic prediction, for their potentialities of application. Genomic prediction was identified as a promising method applicable in oat breeding programs for FHB resistance.

Of particular concern in oat is the asymptomatic infection caused by *Fusarium langsethiae*, a pathogen able to produce the highly toxic T-2 and HT-2 mycotoxins. Isidro-Sánchez et al. (2020a) inoculated 190 spring oat varieties with a mixture of three isolates of the pathogen. Varieties were genotyped using 16,863 genotyping by sequencing markers. GWA identified a single QTL in the linkage group *Mr06* associated with T-2 + HT-2 mycotoxin accumulation. In the QTL, the locus *avgbs_6K_95238.1*encodes lipase precursor that is associated with resistance to fungi. Lipid-derived secondary metabolites produced by the plants are known to play a crucial role in host–pathogen communication. The authors concluded that *Mr06* linkage group plays an important role in *F. langsethiae* resistance.

Willforss et al. (2020) carried out the first proteogenomic study to understand the molecular response of oat when exposed to FHB. The proteomes of resistant and susceptible cultivars were compared, and candidate proteins of interest were identified and linked to protein sequence variants.

6.2 Breeding for Grain Size, Milling, and Naked Grains

Some traits are relevant for processing efficacy and determine quality perception in oat raw materials. A lot of characteristics, such as groat to hull ratio, ease of dehulling, uniformity of groat size, uniformity of mature kernels from top to bottom of the panicle, overall groat size in the medium to large range, few trichomes, husk content, test weight, grain size and weight, and milling yield, are not strictly definable nutritionally important characteristics but from an agronomic and industrial point of view are crucial. Grain yield and kernel size, shape, and morphology have been targets of traditional breeding programs (Fig. 5). Such traits can be positively or negatively linked with health-related characteristics. Negative correlations have been shown among water binding capacity, ß-glucans, intensity of odor, toasted odor, and flavor (Lapveteläinen et al. 2001). High oil concentration was associated with low groat weight and with long and slim seeds (Peterson and Wood 1997). In *A. sterilis,* negative correlations have been found between protein vs. groat size and oil (Rezai and Frey 1988). Selection for high yield and seed weight during oat breeding may have caused a loss of valuable characters health-wise and taste-wise (Lapveteläinen et al. 2001); all the traits can be made available from less developed materials.

Fig. 5 Oat seed biodiversity

Recently, 57 QTLs, influencing one or more of 6 key milling traits, were identified by Klos et al. (2021) in a genome-wide association study involving 501 accessions from the Collaborative Oat Research Enterprise (CORE) panel. The Qkernel QTL was identified as the prominent one, influencing several milling traits. The comparison of QTLs associated with kernel morphology, grain yield, test weight, and chemical groat composition, with those affecting quality-related traits can be important for MAS (Groh et al. 2001). In breeding programs where selection is based on markers linked to specific QTLs, the potential negative associations among characteristics can be a serious limit, suggesting that it is important to have a deeper knowledge of the epistatic impact of specific QTLs.

Naked oats are very interesting materials in the frame of human nutrition; however, few naked cultivars have been developed in comparison to covered types. For example, the naked oats varieties registered in European catalog are one-tenth compared to covered oats. The modern naked varieties are not fully naked but have a percentage of covered kernels, depending on the environment and on the genotype.

The genetics of hulless character is in fact complex. N-1 locus has been identified as the major gene controlling naked trait in oats (Ubert et al. 2017); however, it is incompletely dominant, and the phenotype is the result of the interaction among N-1 gene with modifying genes *N-2, N-3,* and *N-4.* Different level of nakedness is observed depending on the alleles present at each of the four loci. Even the environment can influence this character. The naked character is in fact under genetic control but can be influenced by temperature and humidity.

The yield of naked oats has been traditionally considered lower in comparison to covered varieties. Several explanations have been proposed, for example, the flower morphology (Valentine 1995; Burrows et al. 2001). A high number of varieties was taken into account in the study of Buerstmayr et al. (2007), demonstrating that the yield of covered lines was significantly higher than that of naked lines. Other studies deny the association between naked seeds and lower productivity. Peltonen-Sainio (1997) studied three naked and two covered lines at varying nitrogen fertilization and seeding rates in Finland and found that, under northern growing conditions, groat yield of naked oats was already similar to that of covered oats. Burrows et al. (2001) compared near-isogenic naked and covered lines and showed there was no significant difference in groat yield between the covered and naked isolines. Doehlert et al. (2001) found that the naked oat cultivar Paul produced more than covered oat cultivars over 12 environments during 3 years in the United States. Similar results were obtained by Moudrý et al. (2004), whereas Maunsell et al. (2004) found no significant difference between covered and naked lines for groat production during 2001–2002 in the United Kingdom.

To summarize, naked oats have some disadvantages, such as the lower grain yield and the reduced level of germination, that have limited the use of these cultivars, but have unique end-use advantages, suggesting that specific breeding programs are a crucial step to identify the suitable naked oat genotypes to produce foods of high nutritional value (Redaelli et al. 2015).

6.3 Breeding for β-Glucans

The AVEQ project (Avena Genetic Resources for Quality in Human Nutrition) confirmed the existence of variability for beta-glucan content in a panel of European oats varieties and wild relatives (Redaelli et al. 2013). Although there is genetic variability for β-glucan content and the interest from a nutritional point for this trait, the development of varieties with improved content of β-glucans has been, until now, very slow. Three main reasons slowed down the breeding for this trait: (i) the limits in phenotyping for β-glucan content; (ii) the poorly understood genetics of this complex trait; and (iii) the influence of environment.

From a genetic point of view, the *β*-glucan content is controlled by many loci with additive effects and is also influenced by the environment. The broad sense heritability of this trait has been found ranging from 0.27 to 0.58 (Holthaus et al. 1996; Humphreys and Mather 1996; Kibite and Edney, 1998). *β*-Glucan content is therefore influenced by the environment, but the genotype has a major impact on the expression of this trait (Dvončová et al., 2010). Moreover, genotype–environment interactions have been evaluated, obtaining contrasting indications.

The selection for beta-glucan content is expensive: the most common method for their determination is enzymatic, based on the specific enzymatic degradation of the carbohydrate followed by quantification of the products. In addition, other chemical and physical approaches have been developed, together with fast method of screening, mainly based on NIR.

Because of the complex genetics and the expensive phenotyping, this trait is an ideal candidate for its molecular quantitative trait dissection and selection assisted by molecular markers in breeding programs.

Kianian et al. (2000) suggested that, because of the complexity of the character, a mapping population with a large number of individuals is needed to detect small phenotypic effects. In their work, two recombinant inbred populations with the common parent Kanota were evaluated for beta-glucan content in multiple US environments and two genomic regions hosting the markers Xcdo665B and Xcdo400 were identified able to explain more than 20% of the phenotypic variance of the character in both populations.

Several other studies were focused on the molecular mapping for β-glucan content in oats.

De Koeyer et al. (2004), in a mapping population derived from a hulless spring oat found in North America, detected five QTLs influencing β-glucan content and explaining 23% of the phenotypic variance. Marker cdo484a was identified as associated with the most consistent region.

In the double haploid oat mapping population “Aslak” × “Matilda” (AM), four relevant QTLs were found as responsible for more than one-third of the phenotypic variation for β-glucan content (Tanhuanpää et al. 2012). Two of these QTLs were environment-specific, and two of them are coincident with those identified by Kianian et al. (2000).

Herrmann et al. (2014) identified three loci influencing β-glucan content in two mapping populations grown at three sites in Germany over a 3-year period. The QTL

with the largest effect was QBg1.jki.A-1. This QTL is flanked by the markers E36M55_1 and E36M52_3, explains 31% of the phenotypic variance, and is located on a linkage group putatively homologous to KO_6 of the KO map.

GWA mapping studies have also been published.

Newell et al. (2012), using an oat panel composed of 431 genotypes, identified three markers associated with β-glucan content. One of the markers had high sequence homology to rice chromosome 7 in a region adjacent to CslF gene family members.

Studying 446 oat lines genotyped with 1005 diversity array technology (DArT) markers, Asoro et al. (2013) identified 37 markers explaining β-glucan content. These authors compared efficiency of genomic, marker-assisted, and best linear unbiased prediction (BLUP) for selection for *β*-glucans concentration and concluded that, despite genomic method being more efficient in improving *β*-glucan concentration, it also leads to faster loss of genetic diversity.

Zimmer et al. (2020) evaluated in multiyear trial the beta-glucan content of 413 accessions belonging to the UFRGS Oat Panel, characterized by a weak population structure. The beta-glucan content was associated to seven QTLs located in genomic regions synthenic with barley.

In summary, despite several studies, the genetics underlying the beta-glucan synthesis and regulation in oats is still largely unknown. Important candidate genes for enzymes involved in the beta-glucan synthesis belong to the CLS (cellulose synthase like) gene family. Fogarty et al. (2020), in a genome-wide association study (GWAS) on three panels of elite accessions (spring, winter, world diversity) of oat grown in multiple North American locations, identified 58 significantly associated markers. The homology among the QTLs identified suggested that multiple copies of β-glucan biosynthesis genes are present in the three sub-genomes of oats and contribute to the overall phenotype. In particular, the high expression level of AsCslF6_A and AsCslF6_D genes, located respectively on A and C sub-genomes, is shared by high beta-glucan varieties.

6.4 Breeding for Oil

Oats are richer in oil (~6–10%) than any other cereal, and recurrent selection breeding resulted in lines with up to 18% oil. The oils are mainly concentrated in bran and endosperm and not in the germ, such as other cereals, for example, maize. The localization of the majority of oat grain oil to the endosperm suggests that there is a great potential to increase the oil content of this crop since this storage tissue is the major part of a cereal grain. Oat grains are rich in linoleic acid (Leonova et al. 2008; Leonova, 2013), polar lipids, and peculiar galactolipids with esterified hydroxyl fatty acids (Moreau et al. 2008).

Cultivars with high- or low-oil content have different uses in human and animal nutrition and are both necessary, depending on the destination and on the different technological treatments. For example, low-oil cultivars are generally preferred for rolled oats to simplify the milling procedure and in general are for low-fat foods. On

the contrary, high-oil oats are preferred for feed because more energetic, as well as to produce "oat milk" for human consumption due to the higher nutrition value. Breeding efforts, therefore, must be directed both to the selection of high- and low-oil content oats and even to oats enriched in beneficial PUFA.

Oil content has a high heritability due to additive gene action (Bjørnstad et al. 1994). Negative correlation has been identified between fat and starch contents. Contrasting results have been obtained studying the correlation among fat level, β-glucan, and protein content. It was, in fact, found that the protein concentration increases with nitrogen fertilization and the oil decreases, but there are even cultivars with high protein and oil levels.

An interesting point is the beneficial PUFA 18:2 (linoleic acid) and 18:3 (linolenic acid) content of oats. Leonova et al. (2013) reported that the cultivated oats are richer in such molecules in comparison with wild relatives, even if these last have the highest total oil content. This observation has been confirmed by Dhanda (2011) in an extensive study on 917 oat accessions. The cultivated hexaploidy oats confirmed the elevated levels of linoleic and linolenic acids in different cultivation environments. The environmental impact on the total oil content has been observed, but its influence is still controversial on the oil composition, as reviewed by Stewart and McDougall (2014).

Carlson et al. (2019) used multivariate GWAS methods for identifying genetic associations with metabolites that share lipid biosynthetic pathway. The authors analyzed ten fatty acids simultaneously with multi-trait GWAS methods and found a panel of SNPs that were not detected by traditional, univariate GWAS. They suggested how such multivariate GWAS may increase analysis sensitivity and be advantageous for oat, where the effect of a single locus may be buffered by the activity of its homeologs.

Campbell et al. (2021) developed a trait-specific genomic relationship matrices (TGRMs) model as a prediction tool for the oil content in oats. This model is based on the relationships between individuals using genome-wide markers (SNPs) and place greater emphasis on markers relevant to the trait. In the first step of their work, nine fatty acids were quantified in a panel of 336 genotyped oat lines and the markers effects were used to construct TGRMs. In the second step of the study, the model was validated to predict total seed lipid content in an independent panel of 210 oat lines. The results obtained sustain the utility of using TGRM and improve genomic prediction for a conventional agronomic trait.

The oat oil content is linked with lipase level. Oat lipases are a postharvest problem, which negatively impact on the seed storage, because of the transformation of oils into free fatty acids and the rancid flavor development. In the frame of CropTaylor project, 900 EMS mutation lines were screened for lipase content with the final aim to select low-lipase oat lines (Marmon 2022). Alternatively, or additionally, oat-based antioxidative extracts have been evaluated to prolong oat oil stability. AVA-rich extracts were found to inhibit lipase action and protect toward oxidation. Such natural extracts have therefore the potentialities to be used in oat-based food preservation (Tullberg 2022).

6.5 Breeding for Avenanthramides

Avenanthramides (AVNs) have not been yet a direct target in oat breeding; however, varietal differences in a range from 40 to 132 mg/kg have been found (Dimberg et al., 1993). Moreover, their abundance has been found modulated by the environment (Redaelli et al. 2016). AVN concentration is a heritable trait, as demonstrated by Michels et al. (2020). These authors quantified AVNs in 100 at accessions in a multilocation and multiyear trial, finding that, despite environmental influence, genotype has the largest impact on all three AVN major types of production.

In the review by Shelenga et al. (2021) are reported several screening works in which AVN concentration was measured in several accessions of different origin and species grown in European, Asiatic, and American environments. The highest range of variation for AVN was reported in the study of Leonova et al. (2020), who analyzed the AVN content and composition in 32 wild and 120 cultivated oat accessions, finding a great variability for this trait, with AVN ranging from 4 to 1825 mg kg^{1}.

Brzozowski et al. (2022), in an oat germplasm screening, found, in modern varieties, a reduced genetic variation for these molecules in comparison with obsolete cultivars and landraces. Modern varieties show that the abundance of these molecules increases with the seed size.

Moreover, GWAS and TWAS analyses indicate that the enzymes preceding committed biosynthetic steps seem to have a key role in avenanthramide content. Hernandez-Hernandez et al. (2021) proposed mutagenesis coupled with high-precision biochemical selection as useful method to develop stable lines with a high concentration of total and/or individual AVNs in the oat seed grain.

6.6 Breeding for Protein

Oats are interesting from the point of view of protein content: it is in fact a low-cost protein source and has a higher level of protein in comparison with other grains, including cereals, legumes, and oil seeds. Higher protein is a desirable characteristic of oats used for both food and feed purposes. Moreover, a diet rich in plant protein is gaining attention as it is perceived as healthier for the consumer and more sustainable for the planet in comparison with animal proteins (Kumar et al. 2021).

Proteins in oats are mainly localized in the germ and the bran and less in the endosperm (Boukid 2021). They differ in structural and distributional properties from other cereals that are rich in prolamins and poor in lysine. Globulins are the primary storage protein in oats; they also contain higher concentration of lysine and other essential amino acids, thus making them a superior protein source (Valentine and Cowan 2004).

To improve protein concentration, emphasis has been given to breeding and selections. Sunilkumar et al. (2017) identified, in a mutagenized population, 15 oat

lines with a protein content ranging from 17% to 24%. Such lines, increased in globulin-like proteins, are stable for this character and can be the starting point for the development of a high-quality, high-protein oat variety, both for food and feed. Batalova et al. (2019) worked on naked oats to obtain the high-protein Vitovet cultivar.

7 Are Oats Genetically Modified Crop?

Oats are not genetically engineered crop. No genetically modified oat variety is present on the market. Several reasons are behind this: the complex oat genome is one, but the most important is the almost nonexistent demand to justify the expensive research that goes into developing genetically modified seeds.

Genetically modified oat plants have been obtained with the unique aim to identify suitable protocols for transformation or to study specific gene functions. Oats can be transformed, and to this aim some protocols have been developed for a long time. In the pioneering work of Somers et al. (1992), friable, embryogenic oat callus cultures were obtained following microprojectile bombardment with a plasmid encoding the *Streptomyces hygroscopicus bar* gene and the *Escherichia coli uidA* gene. Pawlowski and Somers (1998) proposed a mechanism of transgenes integration in oat genome at multiple clustered DNA replication forks to explain the observation that all transgenic lines analyzed exhibited genomic interspersion of multiple clustered transgenes. Zhang et al. (1999) transformed shoot meristematic cultures derived from seedlings of commercial varieties.

8 The Oat Genomes

A. sativa is allohexaploid ($2n = 6x = 42$) with a large and complex genome (> 12 Gb), composed of the three AACCDD subgenomes, highly repetitive and characterized by duplications and intra- and intergenomic rearrangements. All these features made full genome assembly a difficult task to reach. An oat reference genome has been therefore missing for decades.

Common oat is thought to have been domesticated from wild-weedy *A. sterilis* L. (Zhou et al. 1999), an allotetraploid carrying CCDD subgenomes and an A_sA_s diploid (Yan et al. 2016). Several diploids carrying A genome variants (A_c, A_d, A_l, A_p, and A_s) (Loskutov and Rines, 2011) are known to be potential source of genes relevant for qualitative trait, for example, improving soluble fiber and protein (Welch et al. 2000).

The C-subgenome carries a putative *CSlF6c* locus that likely has a negative effect on seed soluble fiber content (Coon 2012; Jellen et al. 1994).

Maughan et al. 2019 produced the first, reference-quality, whole-genome reference assemblies for A_s- and C_p-genomes, using *A. atlantica* and *A. eriantha* as plant materials and a hybrid approach for sequencing, involving PacBio long reads, Illumina short reads, and both in vitro and in vivo chromatin-contact mapping.

All these genomic data, together with genetic, germplasm, and phenotypic datasets for oats, are available in GrainGenes (https://wheat.pw.usda.gov or https://graingenes.org), the international centralized repository for curated, peer-reviewed datasets of small grain cereals, including oat. Since 1992, GrainGenes has been a useful source of data and information for oat geneticists and breeders in both public and private sectors worldwide (Blake et al. 2019).

Very recently, the challenge of obtaining a reference genome for *Avena sativa* was also met. PepsiCo announced in 2021 to have completed, in the frame of a public–private collaborative work, the 21 chromosome DNA assembly of the North American oat variety OT3098 using the long-read PacBio technology. The assembly has been made publicly available via the GrainGenes website: the data are hosted on the USDA Agricultural Research Service's GrainGenes website at https://wheat.pw.usda.gov/jb/?data=/ggds/oat-ot3098-pepsico and the datasets can be downloaded at https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico.

In 2022, Kamal et al. (2022) presented the high-quality reference genome of *A. sativa* and of its diploid progenitor *A. longiglumis* and *A. insularis*. The mosaic structure of the oat genome was revealed, characterized by large-scale genomic reorganization during the polyploidization process. The availability of the high-quality reference genome was exploited to clarify some genetic features of health-related traits. For example, the high content and quality of β-glucans are driven in oat by allelic variation and are modulated by transcription factors.

Such studies are the starting point for the future challenge of obtaining oat's pangenome (Mascher 2022). The international PanOat consortium is working on this, with the aim to produce a pan genome deriving from the sequence assemblies of 20 carefully selected oats.

9 Conclusions and Future Prospects

The value of oats in human health is indisputable and therefore is more than rational to focus on this crop through various improvement and enhancement actions. This means new opportunities for using oats through advanced fractionation and the development of innovative nutritious food products. Moreover, oat is a low-input cereal, and its adoption also increases environmental and economic sustainability of cereal-based rotations.

From a policy perspective, the sustainable production of resilient cereal-based crop systems must be combined with systems capable of adapting (to) and mitigating climate change as well as protecting our soils and water courses and maintaining or enhancing biodiversity.

To expand oat cultivation and increase its use, future oat production will need to meet the challenges posed by climate change and the demands of various end users by identifying cultivars with higher and more stable yield, enhanced grain quality, and reduced environmental impacts. This will be complemented by the identification and development of new markets for improved and innovative food, products from oats. Powerful enabling technologies for the identification of specific genes and markers

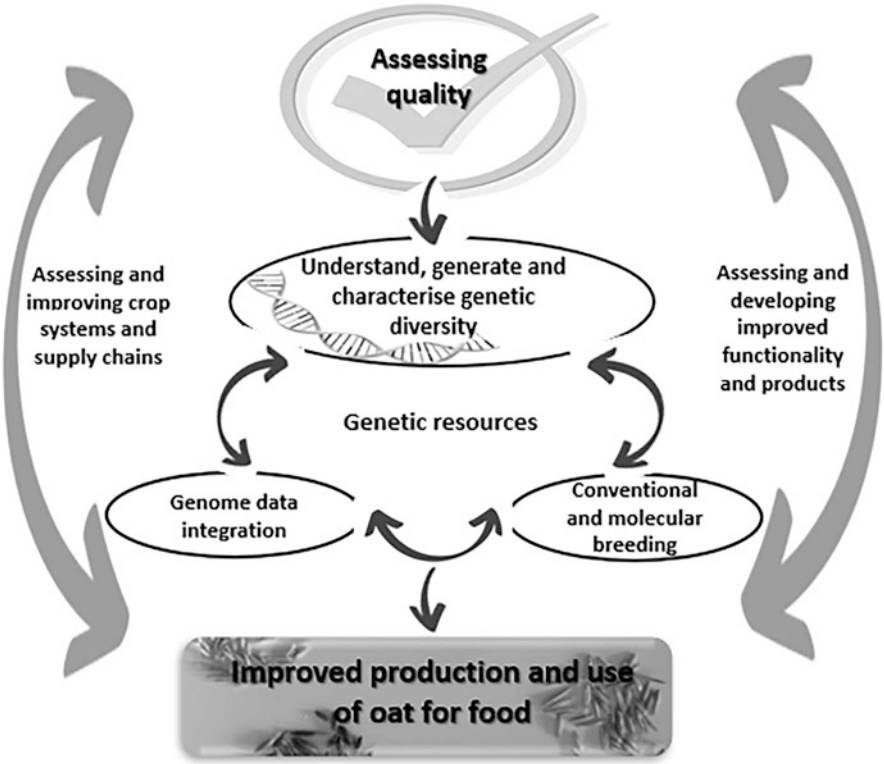


Fig. 6 Interactions aimed at oats valorization for food chains

and accessing genetic diversity will drive the development of breeder-friendly tools accelerating the production of more prolific and resilient oat cultivars. This approach, depicted in Fig. 6, will rely on the use of oat genetic resources, thus enhancing their conservation through plant breeding and deployment of newly bred cultivars with desired traits. An integrated approach is required to bring together academic researchers and applied R&D teams to maximize the translation of research outputs into impact through commercial practice. An example of such interaction is the Scan Oat consortium, whose activities are based on six pillars, ranging from genetics and agronomy to technology and industrial applications, and aimed to "develop food industry using oat as fundamental ingredients" (L. Bülow, https://scanoats.se).

References

Ahmad M, Jehangir IA, Rizvan R, Dar SA, Iqbal S, Wani SH, Mehraj U, Hassan R (2020) Phylogenetic relationship of oats (*Avena sativa* L): a guide to conservation and utilisation of genetic resources. Int J Curr Microbiol App Sci 9(11):831–845. https://doi.org/10.20546/ijcmas.2020.911.101

Asoro FG, Newell MA, Beavis WD, Scott MP, Jannink J-L (2011) Accuracy and training population design for genomic selection on quantitative traits in elite north American oats. Plant Genome 4(2). https://doi.org/10.3835/plantgenome2011.02.0007

Asoro F, Newell MA, Beavis WD, Scott MP, Tinker NA, Jannink J-L (2013) Genomic, marker-assisted, and pedigree-BLUP selection methods for β-glucan concentration in elite oat. Crop Sci 53:1894–1906. https://doi.org/10.2135/cropsci2012.09.0526

Bai J, Ren Y, Li Y, Fan M, Qian H, Wang L, Wu G, Zhang H, Qi X, Xu M, Rao Z (2019) Physiological functionalities and mechanisms of β-glucans. Trends Food Sci Technol 88:57–66. https://doi.org/10.1016/j.tifs.2019.03.023

Barker GLA, Edwards KJ (2009) A genome-wide analysis of single nucleotide polymorphism diversity in the world's major cereal crops. Plant Biotechnol J 7:318–325. https://doi.org/10.1111/j.1467-7652.2009.00412.x

Batalova GA, Loskutov IG, Shevchenko SN, Zhuikova OA, Krotova NV, Tulyakova MV (2019) On breeding of naked oat cultivar virovets. Russ Agric Sci 45(5):412–415. https://doi.org/10.3103/S1068367419050033

Bekele WA, Itaya A, Boyle B, Yan W, Mitchell Fetch J, Tinker NA (2020) A targeted genotyping-by-sequencing tool (Rapture) for genomics-assisted breeding in oat. Theor Appl Genet 133(2):653–664. https://doi.org/10.1007/s00122-019-03496-w

Bjørnstad Å, Aastveit K, Thoresen KS (1994) The potential of high-oil oats under cool temperate conditions. Acta Agric Scand – B Soil Plant Sci 44(4):219–225. https://doi.org/10.1080/09064719409410249

Bjørnstad Å, He X, Tekle S, Klos K, Huang YF et al (2017) Genetic variation and associations involving Fusarium head blight and deoxynivalenol accumulation in cultivated oat (*Avena sativa* L.). Plant Breed 136(5):620–636. https://doi.org/10.1111/pbr.12502

Blake VC, Woodhouse MR, Lazo GR, Odell SG, Wight CP et al (2019) GrainGenes: centralized small grain resources and digital platform for geneticists and breeders. Database (Oxford) 2019:1–7. https://doi.org/10.1093/database/baz065

Bouchard J, Valookaran AF, Aloud BM, Raj P, Malunga LN, Thandapilly SJ, Netticadan T (2022) Impact of oats in the prevention/management of hypertension. Food Chem 321:132198. https://doi.org/10.1016/j.foodchem.2022.132198

Boukid F (2021) Oat proteins as emerging ingredients for food formulation: where we stand? Eur Food Res Technol 247(3):535–544. https://doi.org/10.1007/s00217-020-03661-2

Boz H (2015) Phenolic amides (avenanthramides) in oats – a review. Czech J Food Sci 33(5):399–404. https://doi.org/10.17221/696/2014-CJFS

Brzozowski LJ, Hu H, Campbell MT, Broeckling CD, Caffe M et al (2022) Selection for seed size has uneven effects on specialized metabolite abundance in oat (*Avena sativa* L.). G3-Genes Genom Genet (Bethesda) 12(3):jkab419. https://doi.org/10.1093/g3journal/jkab419

Buerstmayr H, Krenn N, Stephan U, Grausgruber H, Zechner E (2007) Agronomic performance and quality of oat (*Avena sativa* L.) genotypes of worldwide origin produced under Central European growing conditions. Field Crops Res 101(3):343–351. https://doi.org/10.1016/j.fcr.2006.12.011

Burrows VD, Molnar SJ, Tinker NA, Marder T, Butler G, Lybaert A (2001) Groat yield of naked and covered oat. Can J Plant Sci 81(4):727–729. https://doi.org/10.4141/P00-181

Campbell MT, Hu H, Yeats TH, Brzozowski LJ, Caffe-Treml M, Gutiérrez L et al (2021) Improving genomic prediction for seed quality traits in oat (*Avena sativa* L.) using trait-specific relationship matrices. Front Genet 12:643733. https://doi.org/10.3389/fgene.2021.643733

Carlson MO, Montilla-Bascon G, Hoekenga OA, Tinker NA, Poland J et al (2019) Multivariate genome-wide association analyses reveal the genetic basis of seed fatty acid composition in oat (*Avena sativa* L.). G3-Genes Genom Genet 9(9):2963–2975. https://doi.org/10.1534/g3.119.400228

Chang SC, Saldivar RK, Liang PH, Hsieh YSY (2021) Structures, biosynthesis, and physiological functions of (1,3;1,4)-β-D-Glucans. Cells 10(3):510. https://doi.org/10.3390/cells10030510

Chawade A, Sikora P, Bräutigam M, Larsson M, Vivekanand V et al (2010) Development and characterization of an oat TILLING-population and identification of mutations in lignin and β-glucan biosynthesis genes. BMC Plant Biol 10:86. https://doi.org/10.1186/1471-2229-10-86

Chen O, Mah E, Dioum E, Marwaha A, Shanmugam S et al (2021) The role of oat nutrients in the immune system: a narrative review. Nutrients 13(4):1048. https://doi.org/10.3390/nu13041048
Cloetends L (2022) Synergistic effect of OAT polar lipids and oat beta-glucans on postprandial blood glucose. OAT 2022, 11th international oat conference, Perth, Western Australia, 10–13 Oct 2022
Comino I, Moreno MDC, Sousa C (2015) Role of oats in celiac disease. World J Gastroenterol 21 (41):11825–11831. https://doi.org/10.3748/wjg.v21.i41.11825
Commission Regulation (EU) No 432/2012 (2012) Establishing a list of permitted health claims made on foods, other than those referring to the reduction of disease risk and to children's development and health. Official J Eur Union L 136/1
Coon MA (2012) Characterization and variable expression of the CslF6 homologs in oat (*Avena* sp.). Doctoral dissertation, Brigham Young University
de Carvalho MA, Bebeli PJ, Bettencourt E, Costa G, Dias S et al (2013) Cereal landraces genetic resources in worldwide GeneBanks. A review. Agron Sustain Dev 33:177–203. https://doi.org/10.1007/s13593-012-0090-0
De Koeyer DL, Tinker NA, Wight CP, Deyl J, Burrows VD et al (2004) A molecular linkage map with associated QTLs from a hulless × covered spring oat population. Theor Appl Genet 108: 1285–1298. https://doi.org/10.1007/s00122-003-1556-x
Dhanda RK (2011) Fatty acid composition in diverse oat germplasm. Doctoral dissertation, University of Saskatchewan, Canada
Dimberg LH, Theander O, Lingnert H (1993) Avenanthramides – a group of phenolic antioxidants in oats. Cereal Chem 70(6):637–641
Doehlert DC, McMullen MS, Hammond JJ (2001) Genotypic and environmental effects on grain yield and quality of oat grown in North Dakota. Crop Sci 41(4):1066–1072. https://doi.org/10.2135/cropsci2001.4141066x
Dvoncova D, Havrlentova M, Hlinkova A, Hozlar P (2010) Effect of fertilization and variety on the beta-glucan content in the grain of oats. Żywność Nauka Technologia Jakość 17(3):108
Edwards SG, Imathiu SM, Ray RV, Back M, Hare MC (2012) Molecular studies to identify the Fusarium species responsible for HT-2 and T-2 mycotoxins in UK oats. Int J Food Microbiol 156(2):168–175. https://doi.org/10.1016/j.ijfoodmicro.2012.03.020
EFSA (2009) Scientific opinion on the substantiation of health claims related to beta-glucans and maintenance of normal blood cholesterol concentrations (ID 754, 755, 757, 801, 1465, 2934) and maintenance or achievement of a normal body weight (ID 820, 823) pursuant to Article 13(1) of Regulation (EC) No 1924/2006. EFSA J 7(9):1254
EFSA (2010) Scientific opinion on the substantiation of a health claim related to oat beta glucan and lowering blood cholesterol and reduced risk of (coronary) heart disease pursuant to Article 14 of Regulation (EC) No 1924/2006. EFSA J 8(12):1885
EFSA (2011) Scientific opinion on the substantiation of health claims related to beta-glucans from oats and barley and maintenance of normal blood LDL-cholesterol concentrations (ID 1236, 1299), increase in satiety leading to a reduction in energy intake (ID 851, 852), reduction of post-prandial glycaemic responses (ID 821, 824), and "digestive function" (ID 850) pursuant to Article 13(1) of Regulation (EC) No 1924/2006. EFSA J 9(6):2207
Fogarty MC, Smith SM, Sheridan JL, Hu G, Islamovic E et al (2020) Identification of mixed linkage β-glucan quantitative trait loci and evaluation of AsCslF6 homoeologs in hexaploid oat. Crop Sci 60(2):914–933. https://doi.org/10.1002/csc2.20015
Foresman BJ, Oliver RE, Jackson EW, Chao S, Arruda MP, Kolb FL (2016) Genome-wide association mapping of barley yellow dwarf virus tolerance in spring oat (*Avena sativa* L.). PLoS One 11(5):e0155376. https://doi.org/10.1371/journal.pone.0155376
Gavrilova OP, Gagkaeva TY, Orina AS, Markova AS, Kabashov AD, Loskutov IG (2021) Resistance of oat breeding lines to grain contamination with *Fusarium langsethiae* and T-2/HT-2 toxins. Vavilovskii Zhurnal Genet Selektsii 25(7):732–739. https://doi.org/10.18699/VJ21.083
Gilissen LJWJ, Van der Meer IM, Smulders MJM (2016) Why oats are safe and healthy for celiac disease patients. Med Sci 4:21. https://doi.org/10.3390/medsci4040021

Gnanesh BN, McCartney CA, Eckstein PE, Mitchell Fetch JW, Menzies JG, Beattie AD (2015) Genetic analysis and molecular mapping of a seedling crown rust resistance gene in oat. Theor Appl Genet 128(2):247–258. https://doi.org/10.1007/s00122-014-2425-5

González-Barrios P, Bhatta M, Halley M, Sandro P, Gutiérrez L (2021) Speed breeding and early panicle harvest accelerates oat (*Avena sativa* L.) breeding cycles. Crop Sci 61:320–330. https://doi.org/10.1002/csc2.20269

Gorash A, Armonienė R, Mitchell Fetch J, Liatukas Ž, Danytė V (2017) Aspects in oat breeding: nutrition quality, nakedness and disease resistance, challenges and perspectives. Ann Appl Biol 171:281–302. https://doi.org/10.1111/aab.12375

Groh S, Kianian SF, Phillips RL, Rines HW, Stuthman DD, Wesenberg DM, Fulcher RG (2001) Analysis of factors influencing milling yield and their association to other traits by QTL analysis in two hexaploid oat populations. Theor Appl Genet 103:9–18. https://doi.org/10.1007/s001220100579

Haikka H, Manninen O, Hautsalo J, Pietilä L, Jalli M, Veteläinen M (2020) Genome-wide association study and genomic prediction for *Fusarium graminearum* resistance traits in nordic oat (*Avena sativa* L.). Agronomy 10(2):174. https://doi.org/10.3390/agronomy10020174

Hamblin MT, Close TJ, Bhat PR, Chao S, Kling JG et al (2010) Population structure and linkage disequilibrium in U.S. barley germplasm: implications for association mapping. Crop Sci 50(2):556–566. https://doi.org/10.2135/cropsci2009.04.0198

Härröd M, Larsson K (2011) Dispersions of particles of polar lipids and preparations and use thereof; p. WO2011/149416A1

Hautsalo J, Jalli M, Manninen O, Veteläinen M (2018) Evaluation of resistance to *Fusarium graminearum* in oats. Euphytica 214:139. https://doi.org/10.1007/s10681-018-2222-3

Hautsalo J, Jauhiainen L, Hannukkala A, Manninen O, Veteläinen M et al (2020) Resistance to Fusarium head blight in oats based on analyses of multiple field and greenhouse studies. Eur J Plant Pathol 158:15–33. https://doi.org/10.1007/s10658-020-02039-0

He X, Bjørnstad Å (2012) Diversity of north European oat analyzed by SSR, AFLP and DArT markers. Theor Appl Genet 125:57–70. https://doi.org/10.1007/s00122-012-1816-8

He X, Skinnes H, Oliver RE, Jackson EW, Bjørnstad Å (2013) Linkage mapping and identification of QTL affecting deoxynivalenol (DON) content (Fusarium resistance) in oats (*Avena sativa* L.). Theor Appl Genet 126:2655–2670. https://doi.org/10.1007/s00122-013-2163-0

Hernandez-Hernandez O, Pereira-Caro G, Borges G, Crozier A, Olsson O (2021) Characterization and antioxidant activity of avenanthramides from selected oat lines developed by mutagenesis technique. Food Chem 343:128408. https://doi.org/10.1016/j.foodchem.2020.128408

Herrmann MH, Yu J, Beuch S, Weber WE (2014) Quantitative trait loci for quality and agronomic traits in two advanced backcross populations in oat (*Avena sativa* L.). Plant Breed 133(5):588–601. https://doi.org/10.1111/pbr.12188

Holthaus JF, Holland JB, White PJ, Frey KJ (1996) Inheritance of β-glucan content of oat grain. Crop Sci 36:567–572. https://doi.org/10.2135/cropsci1996.0011183X003600030006x

Hossain MM, Tovar J, Cloetens L, Florido MTS, Petersson K, Prothon F, Nilsson A (2021) Oat polar lipids improve cardiometabolic-related markers after breakfast and a subsequent standardized lunch: a randomized crossover study in healthy young adults. Nutrients 13(3):988. https://doi.org/10.3390/nu13030988

Hu X-Z (2022) Processing effects on functional nutrients and digestion characteristics of Chinese OAT products. OAT 2022, 11th international oat conference, Perth, Western Australia, 10–13 Oct 2022

Humphreys DG, Mather DE (1996) Heritability of β-glucan, groat percentage, and crown rust resistance in two oat crosses. Euphytica 91:359–364. https://doi.org/10.1007/BF00033098

Isidro-Sánchez J, D'Arcy Cusack K, Verheecke-Vaessen C, Kahla A, Bekel W et al (2020a) Genome-wide association mapping of *Fusarium langsethiae* infection and mycotoxin accumulation in oat (*Avena sativa* L.). Plant Genome 13:e20023. https://doi.org/10.1002/tpg2.20023

Isidro-Sánchez J, Prats E, Howarth C, Langdon T, Montilla-Bascón G (2020b) Genomic approaches for climate resilience breeding in oats. In: Kole C (ed) Genomic designing of climate-smart cereal crops. Springer, Cham. https://doi.org/10.1007/978-3-319-93381-8_4

Izydorczyk MS, Dexter JE (2008) Barley β-glucans and arabinoxylans: molecular structure, physicochemical properties, and uses in food products – a review. Food Res Int 41(9):850–868. https://doi.org/10.1016/j.foodres.2008.04.001

Jellen EN, Gill BS, Cox TS (1994) Genomic in situ hybridization differentiates between A/D-and C-genome chromatin and detects intergenomic translocations in polyploid oat species (genus *Avena*). Genome 37(4):613–618. https://doi.org/10.1139/g94-087

Joseph S (2022) Inhibition of the intestinal glucose transporter GLUT2 and SGLT1 by OAT derived avenanthramides.OAT 2022, 11th international oat conference, Perth, Western Australia, 10–13 Oct 2022

Joyce S, Kamil A, Fleige L, Gahan CG (2019) The cholesterol-lowering effect of oats and oat beta glucan: modes of action and potential role of bile acids and the microbiome. Front Nutr 6:171. https://doi.org/10.3389/fnut.2019.00171

Kamal N, TsardakasRenhuldt N, Bentzer J et al (2022) The mosaic oat genome gives insights into a uniquely healthy cereal crop. Nature 606:113–119. https://doi.org/10.1038/s41586-022-04732-y

Kianian SF, Phillips RL, Rines HW, Fulcher RG, Webster FH, Stuthman DD (2000) Quantitative trait loci influencing β-glucan content in oat (*Avena sativa*, 2n= 6x= 42). Theor Appl Genet 101:1039–1048. https://doi.org/10.1007/s001220051578

Kibite S, Edney MJ (1998) The inheritance of β-glucan concentration in three oat (*Avena sativa* L.) crosses. Can J Plant Sci 78:245–250. https://doi.org/10.4141/P96-148

Kim IS, Hwang CW, Yang WS, Kim CH (2021a) Multiple antioxidative and bioactive molecules of oats (*Avena sativa* L.) in human health. Antioxidants 10(9):1454. https://doi.org/10.3390/antiox10091454

Kim SJ, Jung CW, Anh NH, Kim SW, Park S, Kwon SW, Lee SJ (2021b) Effects of oats (*Avena sativa* L.) on inflammation: a systematic review and meta-analysis of randomized controlled trials. Front Nutr 8:722866. https://doi.org/10.3389/fnut.2021.722866

Klos KE, Huang YF, Bekele WA, Obert DE, Babiker E et al (2016) Population genomics related to adaptation in elite oat germplasm. Plant Genome 9(2). https://doi.org/10.3835/plantgenome2015.10.0103

Klos KE, Yimer BA, Howarth CJ, McMullen MS, Sorrells ME et al (2021) The genetic architecture of milling quality in spring oat lines of the collaborative oat research enterprise. Foods 10(10):2479. https://doi.org/10.3390/foods10102479

Kosová K, Leišová-Svobodová L, Dvořáček V (2020) Oats as a safe alternative to Triticeae cereals for people suffering from celiac disease? A review. Plant Foods Hum Nutr 75(2):131–141. https://doi.org/10.1007/s11130-020-00800-8

Kouřimská L, Pokhrel K, Božik M, Tilami SK, Horčička P (2021) Fat content and fatty acid profiles of recently registered varieties of naked and hulled oats with and without husks. J Cereal Sci 99:103216. https://doi.org/10.1016/j.jcs.2021.103216

Kumar L, Sehrawat R, Kong Y (2021) Oat proteins: a perspective on functional properties. LWT 152:112307. https://doi.org/10.1016/j.lwt.2021.112307

Lapveteläinen A, Alho-Lehto P, Sinn L, Laukkanen T, Lindman T et al (2001) Relationships of selected physical, chemical, and sensory parameters in oat grain, rolled oats, and cooked oatmeal—a three-year study with eight cultivars. Cereal Chem 78(3):322–329. https://doi.org/10.1094/CCHEM.2001.78.3.322

Leonova S (2013) Lipids in seeds of oat (*Avena* spp.), a potential oil crop. Dissertation, Acta Universitatis Agriculturae Sueciae, 2013:49. ISBN 978-91-576-7837-9

Leonova S, Shelenga T, Hamberg M, Konarev AV, Loskutov I, Carlsson AS (2008) Analysis of oil composition in cultivars and wild species of oat (*Avena* sp.). J Agric Food Chem 56(17):7983–7991. https://doi.org/10.1021/jf800761c

Leonova S, Gnutikov A, Loskutov I, Blinova E, Gustafsson K-E, Olsson O (2020) Diversity of avenanthramide content in wild and cultivated oats. Proc Appl Bot Genet Breed 181:30–47. https://doi.org/10.30901/2227-8834-2020-1-30-47

Li Z, Chen Y, Meesapyodsuk D, Qiu X (2019) The biosynthetic pathway of major avenanthramides in oat. Metabolites 9(8):163. https://doi.org/10.3390/metabo9080163

Liu K, Wise ML (2021) Distributions of nutrients and avenanthramides within oat grain and effects on pearled kernel composition. Food Chem 336:127668. https://doi.org/10.1016/j.foodchem.2020.127668

Loskutov IG (2005) Classification and diversity of the genus *Avena* L. compilers. In: Lipman E, Maggioni L, Knüpffer H, Ellis R, Leggett JM, Kleijer G, Faberova I, Le Blanc A (compilers) Cereal genetic resources in Europe. Report of a Cereals Network, 84–90

Loskutov IG (2008) On evolutionary pathways of *Avena* species. Genet Resour Crop Evol 55:211–220. https://doi.org/10.1007/s10722-007-9229-2

Loskutov IG, Khlestkina EK (2021) Wheat, barley, and oat breeding for health benefit components in grain. Plants 10(1):86. https://doi.org/10.3390/plants10010086

Loskutov IG, Rines HW (2011) Avena. In: Kole C (ed) Wild crop relatives: genomic and breeding resources. Springer, Berlin, Heidelberg, pp 109–183. https://doi.org/10.1007/978-3-642-14228-4_3

Loskutov IG, Blinova EV, Gavrilova OP, Gagkaeva TY (2017) The valuable characteristics and resistance to Fusarium disease of oat genotypes. Russ J Genet Appl Res 7:290–298. https://doi.org/10.1134/S2079059717030108

Marmon S (2022) Towards low lipase OAT. OAT 2022, 11th international oat conference, Perth, Western Australia, 10–13 Oct 2022

Mascher M (2022) Towards a pan-genome for hexaploidy OAT. OAT 2022, 11th international oat conference, Perth, Western Australia, 10–13 Oct 2022

Maughan PJ, Lee R, Walstead R, Vickerstaff RJ, Fogarty MC et al (2019) Genomic insights from the first chromosome-scale assemblies of oat (*Avena* spp.) diploid species. BMC Biol 17:92. https://doi.org/10.1186/s12915-019-0712-y

Maunsell C, Macleod M, Nute G, Wade T (2004) AFENO: avian feed efficiency from naked oats. URL: http://randd.defra.gov.uk/Document.aspx?Document=LS3623_1301_FRP.doc

Meenu M, Zhang Y, Kamboj U, Zhao S, Cao L, He P, Xu B (2021) Rapid determination of β-glucan content of hulled and naked oats using near infrared spectroscopy combined with chemometrics. Foods 11(1):43. https://doi.org/10.3390/foods11010043

Mellers G, Mackay I, Cowan S, Griffiths I, Martinez-Martin P et al (2020) Implementing within-cross genomic prediction to reduce oat breeding costs. Plant Genome 13(1):e20004. https://doi.org/10.1002/tpg2.20004

Menon R, Gonzalez T, Ferruzzi M, Jackson E, Winderl D, Watson J (2016) Oats—from farm to fork. Adv Food Nutr Res 77:1–55. https://doi.org/10.1016/bs.afnr.2015.12.001

Michels DK, Chatham LA, Butts-Wilmsmeyer CJ, Juvik JA, Kolb FL (2020) Variation in avenanthramide content in spring oat over multiple environments. J Cereal Sci 91:102886. https://doi.org/10.1016/j.jcs.2019.102886

Mielniczuk E, Skwaryło-Bednarz B (2020) Fusarium Head Blight, mycotoxins and strategies for their reduction. Agronomy 10(4):509. https://doi.org/10.3390/agronomy10040509

Montilla-Bascón G, Rispail N, Sánchez-Martín J, Rubiales D, Mur LA et al (2015) Genome-wide association study for crown rust (*Puccinia coronata* f. sp. *avenae*) and powdery mildew (*Blumeria graminis* f. sp. *avenae*) resistance in an oat (*Avena sativa*) collection of commercial varieties and landraces. Front Plant Sci 6:103. https://doi.org/10.3389/fpls.2015.00103

Moreau RA, Doehlert DC, Welti R, Isaac G, Roth M, Tamura P, Nuñez A (2008) The identification of mono-, di-, tri-, and tetragalactosyl-diacylglycerols and their natural estolides in oat kernels. Lipids 43(6):533–548. https://doi.org/10.1007/s11745-008-3181-6

Moudrý J, Sterba Z, Moudrý Jr J, Bárta J (2004) The comparison of production ability of naked and husked oats. In: Peltonen-Sainio P, Topi-Hulmi M (eds) Proceedings 7th international oat conference/MTT, Agrifood Research Reports, Jokioinen, Finland

Newell MA, Cook D, Tinker NA, Jannink JL (2011) Population structure and linkage disequilibrium in oat (*Avena sativa* L.): implications for genome-wide association studies. Theor Appl Genet 122:623–632. https://doi.org/10.1007/s00122-010-1474-7

Newell MA, Asoro FG, Scott MP, White PJ, Beavis WD, Jannink JL (2012) Genome-wide association study for oat (*Avena sativa* L.) beta-glucan concentration using germplasm of worldwide origin. Theor Appl Genet 125:1687–1696. https://doi.org/10.1007/s00122-012-1945-0

O'Donoughue LS, Sorrells ME, Tanksley SD, Autrique E, Deynze AV et al (1995) A molecular linkage map of cultivated oat. Genome 38(2):368–380. https://doi.org/10.1139/g95-048

Ociepa T (2019) The oat gene pools–review about the use of wild species in improving cultivated oat. J Cent Eur Agric 20(1):251–261. https://doi.org/10.5513/JCEA01/20.1.2044

Ohlsson L, Rosenquist A, Rehfeld JF, Härröd M (2014) Postprandial effects on plasma lipids and satiety hormones from intake of liposomes made from fractionated oat oil: two randomized crossover studies. Food Nutr Res 58. https://doi.org/10.3402/fnr.v58.24465. PMID: 25317122; PMCID: PMC4189013

Oliver RE, Lazo GR, Lutz JD, Rubenfield MJ, Tinker NA et al (2011) Model SNP development for complex genomes based on hexaploid oat using high-throughput 454 sequencing technology. BMC Genomics 12(1):77. https://doi.org/10.1186/1471-2164-12-77

Oliver RE, Tinker NA, Lazo GR, Chao S, Jellen EN et al (2013) SNP discovery and chromosome anchoring provide the first physically-anchoredhexaploid oat map and reveal synteny with model species. PLoS One 8(3):e58068. https://doi.org/10.1371/journal.pone.0058068

Parry MAJ, Madgwick PJ, Bayon C, Tearall K, Hernandez-Lopez A et al (2009) Mutation discovery for crop improvement. J Exp Bot 60(10):817–2825. https://doi.org/10.1093/jxb/erp189

Pawlowski WP, Somers DA (1998) Transgenic DNA integrated into the oat genome is frequently interspersed by host DNA. Proc Natl Acad Sci U S A 95(21):12106–12110. https://doi.org/10.1073/pnas.95.21.12106

Pellizzaro K, Nava IC, Chao S, Pacheco MT, Federizzi LC (2016) Genetics and identification of markers linked to multiflorous spikelet in hexaploid oat. Crop Breed Appl Biotechnol 16(1):62–70. https://doi.org/10.1590/1984-70332016v16n1a10

Peltonen-Sainio P (1997) Groat yield and plant stand structure of naked and hulled oat under different nitrogen fertilizer and seeding rates. Agron J 89(1):140–147. https://doi.org/10.2134/agronj1997.00021962008900010021x

Perrelli A, Goitre L, Salzano AM, Moglia A, Scaloni A, Retta SF (2018) Biological activities, health benefits, and therapeutic properties of avenanthramides: from skin protection to prevention and treatment of cerebrovascular diseases. Oxid Med Cell Long 2018:6015351. https://doi.org/10.1155/2018/6015351

Perrone G, Ferrara M, Medina A, Pascale M, Magan N (2020) Toxigenic fungi and mycotoxins in a climate change scenario: ecology, genomics, distribution, prediction and prevention of the risk. Microorganisms 8(10):1496. https://doi.org/10.3390/microorganisms8101496

Peterson DM, Wood DF (1997) Composition and structure of high-oil oat. J Cereal Sci 26:121–128. https://doi.org/10.1006/jcrs.1996.0111

Poland JA, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. Plant Genome 5:92–102. https://doi.org/10.3835/plantgenome2012.05.0005

Raguindin PF, Itodo OA, Stoyanov J, Dejanovic GM, Gamba M et al (2021) A systematic review of phytochemicals in oat and buckwheat. Food Chem 338:127982. https://doi.org/10.1016/j.foodchem.2020.127982

Redaelli R, Del Frate V, Bellato S, Terracciano G, Ciccoritti R et al (2013) Genetic and environmental variability in total and soluble β-glucan in European oat genotypes. J Cereal Sci 57(2):193–199. https://doi.org/10.1016/j.jcs.2012.09.003

Redaelli R, Scalfati G, Ciccoritti R, Cacciatori P, De Stefanis E, Sgrulletta D (2015) Effects of genetic and agronomic factors on grain composition in oats. Cereal Res Commun 43(1):144–154. https://doi.org/10.1556/CRC.2014.0019

Redaelli R, Dimberg L, Germeier CU, Berardo N, Locatelli S, Guerrini L (2016) Variability of tocopherols, tocotrienols and avenanthramides contents in European oat germplasm. Euphytica 207(2):273–292. https://doi.org/10.1007/s10681-015-1535-8

Ren G, Fan X, Teng C, Li Y, Everaert N, Blecker C (2021) The beneficial effect of coarse cereals on chronic diseases through regulating gut microbiota. Foods 10(11):2891. https://doi.org/10.3390/foods10112891

Rezai A, Frey KJ (1988) Variation in relation to geographical distribution of wild oats-seed traits. Euphytica 39:113–118. https://doi.org/10.1007/BF00039862

Ryan L, Thondre PS, Henry CJK (2011) Oat-based breakfast cereals are a rich source of polyphenols and high in antioxidant potential. J Food Compos Anal 24(7):929–934. https://doi.org/10.1016/j.jfca.2011.02.002

Saeedi-Boroujeni A, Mahmoudian-Sani MR, Nashibi R, Houshmandfar S, TahmasebyGandomkari S, Khodadadi A (2021) Tranilast: a potential anti-inflammatory and NLRP3 inflammasome inhibitor drug for COVID-19. Immunopharmaco Immunotoxicol 43(3):247–258. https://doi.org/10.1080/08923973.2021.1925293

Sánchez-Martín J, Rubiales D, Flores F, Emeran AA, Shtaya MJY et al (2014) Adaptation of oat (*Avena sativa*) cultivars to autumn sowings in Mediterranean environments. Field Crops Res 156:111–122. https://doi.org/10.1016/j.fcr.2013.10.018

Shelenga TV, Kerv YA, Perchuk IN, Solovyeva AE, Khlestkina EK et al (2021) The potential of small grains crops in enhancing biofortification breeding strategies for human health benefit. Agronomy 11(7):1420. https://doi.org/10.3390/agronomy11071420

Shi J, Arunasalam K, Yeung D, Kakuda Y, Mittal G, Jiang Y (2004) Saponins from edible legumes: chemistry, processing, and health benefits. J Med Food 7(1):67–78. https://doi.org/10.1089/109662004322984734

Singh R, De S, Belkheir A (2013) Avena sativa (Oat), a potential neutraceutical and therapeutic agent: an overview. Crit Rev Food Sci Nutr 53(2):126–144. https://doi.org/10.1080/10408398.2010.526725

Smulders MJM, van de Wiel CCM, van den Broeck HC, van der Meer IM, Israel-Hoevelaken TPM et al (2018) Oats in healthy gluten-free and regular diets: a perspective. Food Res Int 110:3–10. https://doi.org/10.1016/j.foodres.2017.11.031

Somers DA, Rines HW, Gu W, Kaeppler HF, Bushnell WR (1992) Fertile, transgenic oat plants. Nat Biotechnol 10(12):1589–1594. https://doi.org/10.1038/nbt1292-1589

Song G, Huo P, Wu B, Zhang Z (2015) A genetic linkage map of hexaploid naked oat constructed with SSR markers. Crop J 3(4):353–357. https://doi.org/10.1016/j.cj.2015.01.005

Soycan G, Schär MY, Kristek A, Boberska J, Alsharif SN (2019) Composition and content of phenolic acids and avenanthramides in commercial oat products: are oats an important polyphenol source for consumers? Food Chem X 3:100047. https://doi.org/10.1016/j.fochx.2019.100047

Stewart D, McDougall G (2014) Oat agriculture, cultivation and breeding targets: implications for human nutrition and health. Brit J Nutr 112(S2):S50–S57

Sun NX, Tong LT, Liang TT, Wang LL, Liu LY et al (2019) Effect of oat and tartary buckwheat–based food on cholesterol–lowering and gut microbiota in hypercholesterolemic hamsters. J Oleo Sci 68(3):251–259. https://doi.org/10.5650/jos.ess18221

Sunilkumar BA, Leonova S, Öste R, Olsson O (2017) Identification and characterization of high protein oat lines from a mutagenized oat population. J Cereal Sci 75:100–107. https://doi.org/10.1016/j.jcs.2017.03.003

Tang Y, Li S, Yan J, Peng Y, Weng W et al (2022) Bioactive components and health functions of oat. Food Rev Int 1–20. https://doi.org/10.1080/87559129.2022.2029477

Tanhuanpää P, Kalendar R, Schulman AH, Kiviharju E (2008) The first doubled haploid linkage map for cultivated oat. Genome 51(8):560–569. https://doi.org/10.1139/G08-040

Tanhuanpää P, Manninen O, Beattie A, Eckstein P, Scoles G, Rossnagel B, Kiviharju E (2012) An updated doubled haploid oat linkage map and QTL mapping of agronomic and grain quality traits from Canadian field trials. Genome 55(4):289–301. https://doi.org/10.1139/g2012-017

Tekle S, Lillemo M, Skinnes H, Reitan L, Buraas T, Bjørnstad Å (2018) Screening of oat accessions for Fusarium Head Blight resistance using spawn-inoculated field experiments. Crop Sci 58:1–9. https://doi.org/10.2135/cropsci2017.04.0264

Tekle S, StråbøSchofer S, He X, Dong Y, Bjørnstad Å (2020) Variation in anther extrusion and its impact on Fusarium Head Blight and deoxynivalenol content in oat (*Avena sativa* L.). Agronomy 10(3):354. https://doi.org/10.3390/agronomy10030354

Tinker NA, Kilian A, Wight CP, Heller-Uszynska K, Wenzl P et al (2009) New DArT markers for oat provide enhanced map coverage and global germplasm characterization. BMC Genomics 10:39. https://doi.org/10.1186/1471-2164-10-39

Tinker NA, Chao S, Lazo GR, Oliver RE, Huang YF et al (2014) A SNP genotyping array for hexaploid oat. Plant Genome 7(3):plantgenome2014-03. https://doi.org/10.3835/plantgenome2014.03.0010

Tripathi V, Mohd AS, Ashraf T (2018) Avenanthramides of oats: medicinal importance and future perspectives. Pharmacogn Rev 12(23):66–71. https://doi.org/10.4103/phrev.phrev_34_17

Tullberg C (2022) Antioxidative potential of natural OAT extracts. OAT 2022, 11th international oat conference, Perth, Western Australia, 10–13 Oct 2022

Tumino G, Voorrip RE, Rizza F, Badeck FW, Morcia C et al (2016) Population structure and genome-wide association analysis for frost tolerance in oat using continuous SNP array signal intensity ratios. Theor Appl Genet 129(9):1711–1724. https://doi.org/10.1007/s00122-016-2734-y

Tumino G, Voorrips RE, Morcia C, Ghizzoni R, Germeier CU et al (2017) Genome-wide association analysis for lodging tolerance and plant height in a diverse European hexaploid oat collection. Euphytica 213:163. https://doi.org/10.1007/s10681-017-1939-8

Tye-Din J (2022) Getting oats on the menu for people with coeliac disease. OAT 2022, 11th international oat conference, Perth, Western Australia, 10–13 Oct 2022

Ubert IP, Zimmer C, Pellizzaro K, Federizzi LC, Nava IC (2017) Genetics and molecular mapping of the naked grains in hexaploid oat. Euphytica 213(2):1–16. Euphytica 213:41. https://doi.org/10.1007/s10681-017-1836-1

Valentine J (1995) Naked oats. In: Welch RW (ed) The oat crop, World crop series. Springer, Dordrecht. https://doi.org/10.1007/978-94-011-0015-1_16

Valentine J, Cowan S (2004) New directions in breeding for high quality oats. In: Peltonen-Sainio P, Topi-Hulmi M (eds.) Proceedings 7th international oat conference/MTT, Agrifood Research Reports, Jokioinen, Finland

Vivekanand V, Chawade A, Larsson M, Larsson A, Olsson O (2014) Identification and qualitative characterization of high and low lignin lines from an oat TILLING population. Ind Crop Prod 59:1–8. https://doi.org/10.1016/j.indcrop.2014.04.019

Wang Q, Ellis PR (2014) Oat β-glucan: physico-chemical characteristics in relation to its blood-glucose and cholesterol-lowering properties. Brit J Nutr 12(S2):S4–S13. https://doi.org/10.1017/S0007114514002256

Welch RW, Brown JCW, Leggett JM (2000) Interspecific and intraspecific variation in grain and groat characteristics of wild oat (avena) species: very high groat (1→ 3),(1→ 4)-β-D-glucan in an *Avena atlantica* genotype. J Cereal Sci 31(3):273–279. https://doi.org/10.1006/jcrs.2000.0301

Willforss J, Leonova S, Tillander J, Andreasson E, Marttila S et al (2020) Interactive proteogenomic exploration of response to Fusarium head blight in oat varieties with different resistance. J Proteome 218:103688. https://doi.org/10.1016/j.jprot.2020.103688

Winkler LR, Michael Bonman J, Cha S, AdmassuYimer B, Bockelman H, EsveltKlos K (2016) Population structure and genotype–phenotype associations in a collection of oat landraces and historic cultivars. Front Plant Sci 7:1077. https://doi.org/10.3389/fpls.2016.01077

Yan W, Fregeau-Reid J, Rioux S, Pageau D, Xue A et al (2010) Response of oat genotypes to Fusarium Head Blight in Eastern Canada. Crop Sci 50(1):134–142. https://doi.org/10.2135/cropsci2009.03.0129

Yan H, Bekele WA, Wight CP, Peng Y, Langdon T (2016) High-density marker profiling confirms ancestral genomes of avena species and identifies D-genome chromosomes of hexaploid oat. Theor Appl Genet 129:2133–2149. https://doi.org/10.1007/s00122-016-2762-7

Yang J, Wang P, Wu W, Zhao Y, Idehen E, Sang S (2016) Steroidal saponins in oat bran. J Agric Food Chem 64(7):1549–1556. https://doi.org/10.1021/acs.jafc.5b06071

Zhang S, Cho MJ, Koprek T, Yun R, Bregitzer P, Lemaux PG (1999) Genetic transformation of commercial cultivars of oat (Avena sativa L.) and barley (Hordeum vulgare L.) using in vitro shoot meristematic cultures derived from germinated seedlings. Plant Cell Rep 18:959–966. https://doi.org/10.1007/s002990050691

Zhang K, Dong R, Hu X, Ren C, Li Y (2021) Oat-based foods: chemical constituents, glycemic index, and the effect of processing. Foods 10(6):1304. https://doi.org/10.3390/foods10061304

Zhou X, Jellen EN, Murphy JP (1999) Progenitor germplasm of domisticatedhexaploid oat. Crop Sci 39(4):1208–1214. https://doi.org/10.2135/cropsci1999.0011183X003900040042x

Zimmer CM, McNish IG, Klos KE, Oro T, Arruda K et al (2020) Genome-wide association for β-glucan content, population structure, and linkage disequilibrium in elite oat germplasm adapted to subtropical environments. Mol Breed 40:103. https://doi.org/10.1007/s11032-020-01182-0

Zohary D, Hop M, Weis E (2012) Domestication of plants in the Old World, 4th edn. Oxford University Press, Oxford

Genetic Improvement of Sorghum: Crop Genome Designing for Nutraceuticals

M. T. Labuschagne and L. Elkonin

Contents

Abstract

Sorghum is an ancient cereal crop grown widely in the dry regions of Africa and Asia, mainly by subsistence farmers. It is inherently adapted to drought and heat stress and has significant potential as a global food security crop in changing

M. T. Labuschagne (✉)
Department of Plant Sciences, University of the Free State, Bloemfontein, South Africa
e-mail: labuscm@ufs.ac.za

L. Elkonin
Department of Biotechnology, Federal Centre of Agriculture Research of the South-East Region, Saratov, Russia

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_6

climatic conditions. The nutritional profile of sorghum is comparable with that of other cereals, but is unique in that it has various bioactive compounds such as phenolic acids, procyanidins, flavonoids, and anthocyanins. The protein (kafirin) of sorghum is, however, poorly digestible, and genetic improvement of protein digestibility has remained a challenge. Large genetic variability has been shown for almost all the bioactive compounds in tested sorghum accessions, but there is a large gap in the knowledge of the genetics underlying their expression. There are a number of sorghum germplasm collections, reference genome sequences, association panels, and mutant populations available in the world, which could be used to screen for genetic variation and determine genetic architecture, in order to improve sorghum nutraceutical content. Genetic engineering techniques have been applied to improve kafirin digestibility and carotenoid content, and have significant potential for improving nutraceutical content in the future. Information on the genes linked to nutraceuticals and the gene action involved in their expression will allow improvement of nutraceutical content through conventional breeding and by use of molecular markers, genomics, and genetic engineering. All available information on grain nutraceuticals and other nutritional components in sorghum should be integrated as a resource to be used by the sorghum community for sorghum improvement and genetic studies of nutraceuticals.

Keywords

Sorghum · Phenolic compounds · Kafirin · Genetic resources · Genomics · Genetic engineering

1 Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is a stress-tolerant C4 plant, which belongs to the *Poaceae* family, which is well adapted to arid growing conditions. It is a versatile ancient cereal grown mainly by subsistence farmers. This annual crop is mostly photoperiod insensitive and completes its life cycle in 4 months. It is seen as a potential food security crop highly suited to climate change conditions. Maize, rice, and wheat are the most important cereals in the world, with sorghum ranking fifth in terms of importance. It is a staple food to more than 500 million people in Africa and Asia (Xin et al. 2021). The annual sorghum production is over 57 million tons (FAOSTAT 2019). In developed countries, it is mainly used for animal feed, biofuel production, and bioproducts, but it is a staple food in many developing countries (Xin et al. 2021).

Nutraceuticals are, by definition, bioactive compounds, which are biologically active, and are found in food. Various phytochemical substances are regarded as bioactive compounds, including the polyphenols. Dietary fiber, polyunsaturated fatty acids, micro- and macronutrients, vitamins, oligosaccharides, lactic acid bacteria, choline, and lecithin are all bioactive compounds (Chaudhari et al. 2017). Biopeptides, which are biologically active peptides, are also nutraceuticals.

Nutraceuticals have possible antioxidant properties and various other effects on the human body. Protein is a natural source of biopeptides, including antioxidant biopeptides (Hu et al. 2021). Certain amino acids, peptides, and proteins are also possible bioactive compounds (Szerszunowicz and Kłobukowski 2020). Reactive oxygen species (ROS) is important in metabolism and aging of humans, and some nutraceuticals contribute to decreasing the levels of ROS formed in the body through antioxidant activity (Li et al. 2021a). Human consumption of foods containing nutraceuticals have many therapeutic advantages. These compounds contribute to longevity, delaying aging, and preventing the development chronic diseases such as diabetes and hypertension, among others (Szerszunowicz and Kłobukowski 2020). Nutraceuticals can contribute to the prevention and management of viral infections (Li et al. 2021a). Different kinds of biological properties and activities have been ascribed to bioactive compounds, including antioxidant, anti-inflammatory, and antimicrobial properties, which can contribute to protection against human disease (Omrani et al. 2020).

Sorghum grains consist of carbohydrates, kafirin (protein), fiber, polyunsaturated fatty acids, and resistant starch (Khalid et al. 2022). Sorghum has a nutritional profile comparable with other cereals, but is unique to other cereals in that it contains different bioactive compounds such as anthocyanins, flavonoids, phenolic acids, and procyanidins (Ofosu et al. 2021). The extensive biological activities of sorghum grains have been shown in a number of in vitro and in vivo studies. Sorghum bioactive compounds have been shown to inhibit oxidative stress, cardiovascular disease, high lipid levels, and hypertension. It has anticancer and antidiabetic properties and could lower cholesterol index and obesity through antioxidant and anti-inflammatory mechanisms (Li et al. 2021a). The phenolic compounds having bioactive properties include phenolic acid, flavonoids, stilbenes, and tannins. Sorghum grain also contains B-complex vitamins, fat-soluble vitamins A, D, E, and K, as well as minerals, including magnesium, potassium, phosphorus, and zinc. Some sorghum polyphenols, including tannin (proanthocyanidin) and 3-deoxyanthocyanidin, have the potential to protect individuals against inflammation, diabetes, and oxidative stress. Flavonoids are also known as antioxidants, which could positively influence inflammatory and neurodegenerative diseases, as well as cancer and diabetes. Sorghum grain also have a high fiber content, which can reduce blood cholesterol and glucose levels (Khalid et al. 2022). As sorghum is a good source of B vitamins, minerals, carbohydrates, and is also gluten free, it has significant potential as a source of food and beverages and gluten-containing replacement diets for people with celiac disease (Bouargalne et al. 2022).

Although sorghum is known for various phytochemicals, which contribute to human health, it is often not valued for these attributes due to the poor digestibility of its proteins (Duressa et al. 2018). Extensive research has been done by various research groups on sorghum protein digestibility. There have been sustained (although limited) global research efforts focusing on unraveling the biochemical and genetic basis of low protein digestibility and the improvement of kafirin bioavailability. The sequencing of the sorghum genome has supported these efforts, leading to a better scientific understanding of kafirin chemical properties, the

genetics underlying kafirins, and its protein body structure. Despite this, there have been several challenges in the genetic improvement of protein digestibility, as increased protein digestibility is linked to agronomically undesirable traits such as the opaque (soft) endosperm phenotype. It is technically possible to develop high-protein digestible sorghum with hard endosperm, as was the case with quality protein maize, especially with the help of marker-assisted breeding (Duressa et al. 2018). Li et al. (2021a) stated that kafirins are some of the best nutraceutical sources.

Significant genetic variability has been reported for almost all the bioactive compounds in numerous sorghum accessions, indicating the possibility for genetic improvement through selection. Despite this, very limited research has been done on the genetic basis of bioactive compounds in sorghum and its possible improvement through conventional breeding, the use of DNA markers, genomics, or genetic engineering. To improve sorghum, it is critical to use the available genetic variation. These genetic resources should be screened for valuable genes underlying the synthesis of nutraceutical compounds (Bouargalne et al. 2022).

In the world, there are a number of sorghum germplasm collections, reference genome sequences of good quality, and association panels, which can be used for genome-wide association studies for food and quality related traits. There are also mutant populations, which are useful to discover genes, which can be applied for sorghum improvement, as well as information on gene expression. Genetic engineering is becoming increasingly important in sorghum nutritional content research, and the technology has developed significantly in recent years.

The aim of this chapter was to determine the current status of sorghum as a source of nutraceuticals, to assess current genetic resources in sorghum for possible genetic manipulation and improvement of nutraceutical content, and the use of new technologies such as genomics and genetic engineering to improve sorghum as a source of nutraceuticals in the human diet.

2 Sorghum Grain Chemical Composition

Sorghum grain composition is similar to that of maize and millet. It consists of starch, lipids, protein, as well as non-starch polysaccharides. It also contains B vitamins and vitamins D, E, and K (fat-soluble), as well as minerals (Przybylska et al. 2019). Sorghum seed was reported to consist of protein (4.4–21.1%), fat (2.1–7.6%), crude fiber (1.0–3.4%), total carbohydrate (57.0–80.6%), starch (55.6–75.2%), and total minerals as ash (1.3–3.5%) (Cabrera et al. 2020). In contrast to other cereals, sorghum is also a rich source of phenolic compounds.

2.1 Phenolic Compounds

Sorghum has the highest amount of phenolic compounds of all cereals. A variety of phenolic compounds are present in sorghum grains, including phenolic acids, flavonols, 3-deoxyanthocyanidins, flavanones, flavones, and condensed tannins

(Shen et al. 2018). The outer layer of the grain (bran) has a high concentration of most of these compounds, which is unfortunately often removed during milling by decortication, which leads to significant reduction of sorghum health benefits. Therefore, the consumption of whole grain sorghum has many key health benefits, such as free radical scavenging activity (Kumari et al. 2021).

Two main categories of phenolic compounds have been identified, namely soluble and insoluble. The soluble compounds like flavonoids, quinones, and phenylpropanoids are found in the vacuole of plant cells, while the insoluble compounds like lignins, condensed tannins, and hydroxycinnamic acid are attached to the cell wall (Gharaati 2019). The phenolic compounds are important secondary metabolites and are major contributors to the antioxidant properties of sorghum, having significant physiological benefits for humans. For this reason, the consumption of sorghum-based foods with high levels of polyphenolic substances can contribute to the prevention and reduction of the risk of chronic diseases, including some cancers and diabetes. The sorghum variety or genotype, the grain pericarp color, and the testa pigmentation all influence the phenolic compound profiles in sorghum. There are four classes of pericarp color in sorghum, which are white, yellow, red, and black. Genetically, black sorghums are actually red, as sunlight during the maturation process causes the red color to turn black. The white sorghums, also known as food-type sorghum, have very low levels of total extractable phenol and usually have no tannins or anthocyanins. Red sorghum is caused by a red pericarp and contains significant levels of extractable phenols, although they have no tannins. Black pericarps lead to black sorghums, which have very high levels of anthocyanins, while the brown sorghums have varying levels of phenolic compounds and pigmented pericarps (Khalid et al. 2022).

The presence of total phenolic compounds in most sorghum whole grain is 0.46 ~ 20 mg gallic acid equivalent (GAE) per gram. The highest total phenolic compound content of almost 48 mg GAE/g was reported in whole grain red sorghum. Total phenolic content in sorghum bran varies even more than in seed. The total phenolic content of red sorghum bran was 20-fold that of white sorghum bran and 8.9-fold that of yellow sorghum bran, while that in black sorghum bran extracts was 7.5-fold that of white sorghum bran and 3.3-fold that of yellow sorghum bran extract (Burdette et al. 2010). It must be kept in mind that the variations in the quantities of total phenolic compounds reported in different studies could have been influenced by different extraction procedures and the solvents used, the specific genotypes tested, and environmental conditions in which plants were grown (Li et al. 2021a).

Free and bound forms of phenolic compounds are found in sorghum, but 70–95% of phenolic acids are in a bound form. Despite this, far more research has been done on the identification and biological activity of free phenolic compounds in sorghum than on the bound compounds (Li et al. 2021b). Bound phenolics bind to structural compounds of the cell wall, thereby reducing bioavailability. It is therefore important to find novel ways to promote the release of bound phenolic compounds in order to increase their bioavailability. The molecular bonding can be severed by creating an acidic or basic environment and by elevating the temperature, which can cause release (Li et al. 2021a; Espitia-Hernández et al. 2022).

Phenolic compounds can act as natural antioxidants by decreasing the oxidative damage of biomolecules, which reduces the effects of reactive oxidants. Phenolic extracts from sorghum were shown to decrease and inhibit the growth of cancer cells in organs such as the colon, liver, and esophagus (Kadri et al. 2017). Phenolic compounds vary significantly between sorghum seeds of different colors (black, brown, red, and white). The amount of phenolic compounds is highly correlated with antioxidant properties. Due to their phenolic compound content, sorghum grains are important as an ingestible form of antioxidants in the diet. The development of sorghum varieties with high levels of antioxidants can increase its nutritional value and health benefits. Sorghum tannins also have medicinal properties (Choi et al. 2019).

2.1.1 Phenolic Acids and Flavonoids

Various studies have reported different types and numbers of phenolic acids, but the most frequently reported phenolic acids isolated from sorghum were caffeic acid and ferulic acid, 3-deoxyanthocyanidins (luteolinidin and apigeninidin), flavanones (naringenin), flavones (luteolin and apigenin), and dihydroflavonol (taxifolin) (Wu et al. 2017). Caffeic acid, p-coumaric acid, sinapic acid, gallic acid, protocatechuic acid, and p-hydroxybenzoic acid have also been reported quite frequently. Ferulic acid is the predominant phenolic acid, especially in red and brown sorghum, and is uncountable for about 90% of the combined phenolic acids. In red sorghum, ferulic acid was the predominant phenolic acid, and p-coumaric, caffeic, and 3,4-dihydroxybenzoic acids were also identified, but there was a large genotype influence on phenolic acids (Li et al. 2021b). Red and brown sorghum grains had the most luteolinidin and apigeninidin, followed by black grains, while white pericarp varieties had very low amounts of these compounds. Many phenolic acids have been assayed, and of these, ferulic, p-coumaric, and protocatechuic acids were seen in the highest concentrations in both red and white sorghum grain (Przybylska-Balcerek et al. 2018).

The outer layers of the grain contain most of the flavonoids, which contribute to the coloring of the pericarp. Sorghum grains have been found to contain many flavonoids. As in the case of phenolic acids, a large genotype effect was seen (Li et al. 2021b). Both the color and thickness of the pericarp influence flavonoid concentrations and profiles. Flavonoids have antioxidant properties, and the daily consumption of foods with significant amounts of flavonoids can help to reduce the risk of cancers of the breast, colon, and pancreas. The main flavonoids in sorghum are anthocyanins, which are a group of anthocyanidin glycosides (Wu et al. 2017). Anthocyanin content in sorghum bran was three to four times higher than that of the flour. The highest anthocyanin content was reported in black sorghum bran (Kumari et al. 2021).

The most frequently reported flavones in sorghum grains are luteolin, apigenin, and naringenin. The widest known flavonols in sorghum grain are kaempferol and quercetin, and the most researched flavanol is iscatechin, and for dihydroflavonol, it is taxifolin (Luo et al. 2020). Ofosu et al. (2021) were the first to report the presence of formononetin, glycitein, and ononin in sorghum. High concentrations of flavanones were identified in yellow-pigmented sorghum genotypes, and those with colored testas had a higher content of condensed tannins (Szerszunowicz and

Kłobukowski 2020). A high concentration of flavones was reported in red-brown sorghum bran (576.47 μg/g) and lower amounts in yellow sorghum flour (15.3 μg/g). The flavanone concentrations were higher in yellow bran (1773.47 μg/g) and lower in brown sorghum flour (4.29 μg/g). Flavanone concentrations were consistently higher in bran than in flour (Kumari et al. 2021). Some of the sorghum flavones had estrogenic effects, as well as anticancer effects in vitro (Cox et al. 2019).

Of all the cereals, only sorghum is a dietary source of 3-deoxyanthocyanidins (3-DXAs). These occur mainly as luteolinidin and apigeninidin, which are mostly water-soluble pigments (Luo et al. 2020). They produce yellow (apigeninidin) and orange (luteolinidin) colors in acidic solvents. They are effective natural colorants and have good antioxidant properties, which are beneficial for human health. Li et al. (2021a) reported that luteolinidin was the predominant 3-DXA, with its total content accounting for 40.55–78.36% of the total 3-DX. The concentration of 3-DXA was three to four times higher in black testa seeds than red or brown testa seeds (Shen et al. 2018). Kumari et al. (2021) reported that red-brown sorghum bran had the highest 3-DXA (4479.16 μg/g), while yellow sorghum flour had low 3-DXA levels (14.14 μg/g). Results from in vitro tests in sorghum showed that 3-DXA had both anticancer and antioxidant properties (Cox et al. 2019). The main antidiabetic substances in sorghum flavonoids have likewise been ascribed to condensed tannins and 3-DXAs (Li et al. 2021a).

2.1.2 Stilbenoids

Stilbenoids are a group of phenolic compounds, and they have a number of benefits for human health. Sorghum can produce stilbenoid metabolites, but very limited research has been done on this. The total stilbenoid content is related to grain color. One study reported 0.4–1 mg/kg of trans-piceid in white sorghum and up to 0.2 mg/kg trans-resveratrol in red sorghum grains (Bröhan et al. 2011).

2.1.3 Tannins

The growing knowledge on the health benefits of tannins have led to tannins receiving more attention in the last years. They are heterogeneous polyphenolic polymers. Tannins in sorghum occur mainly in the pericarp and are polymerized products of flavan-3, 4-diols, and/or flavan-3-ols. Tannin content in sorghum bran is higher than in the flour. Bran from brown and red-brown sorghum varieties had significantly higher tannin content than bran of red and yellow varieties (Kumari et al. 2021). Tannins have many health benefits as they have antioxidant properties and are radical scavengers. They were also reported to improve immunity and have anticancer and anti-inflammatory properties. They are cardioprotective, are vasodilators and have antithrombotic effects (Queiroz et al. 2018). Sorghum is an excellent source of tannins compared to other cereals. Seyoum et al. (2016) reported tannin concentration ranging from 0.2 to 48.0 mg/g in sorghum, with grain with a black testa having the highest tannin level. A significant growing season effect was evident on the tannin content and activity in sorghum. Sorghum with a high tannin content generally has good resistance to birds, insects, and molds.

Tannins, unfortunately, also have anti-nutritional effects, as they form complexes with protein and iron. This impairs the digestibility of the protein and reduces iron absorption (Iyabo et al. 2018). Apart from their interaction with protein, tannins also affect carbohydrates, especially starch, hemicellulose, cellulose, and pectin, reducing their digestibility. Based on their biological and chemical characteristics, tannins have been grouped into hydrolyzable and condensed (known as proanthocyanidins) tannins. Hydrolyzable tannins are complex polymeric compounds classified as gallotannins, when derived from gallic acid, and ellagitannins, when derived from ellagic acid, which is a dimer of gallic acid. Many and varied hydrolyzable tannins have been isolated from edible and inedible plants. They have anticancer, antidiabetic, and antibiotic effects. Condensed tannins are the main polyphenolic substances in sorghum (De Oliveira et al. 2017).

Condensed tannins are especially common in sorghum seeds with pigmented testas. The testa is a structure present between the pericarp and the endosperm of the grain of only some varieties. Sorghum varieties with a brown testa generally have more procyanidins than that of other seed colors. The presence of condensed tannins is one of the reasons why sorghum has higher levels of antioxidants than any other cereal (De Oliveira et al. 2017). Proanthocyanidins in sorghum grains have antioxidant, antitumor, and lipid-lowering activities, and they play a role in the prevention of cardiovascular diseases. They are powerful radical scavengers (Yu et al. 2018). The most basic unit of composition of condensed tannins is catechin, gallocatechin, allocatechin gallate, and afzelechin (Jiang et al. 2020). Testa color is used a means to classify sorghum grains into three categories where type I is tannin-free sorghum without pigmented testa, type II sorghum has a pigmented testa layer (condensed tannins), and type III contain tannins in the testa and pericarp. The proanthocyanidins in sorghum form complexes and precipitate with proteins, probably causing both acidity and a bitter taste, which contribute to them being bird repellents (Li et al. 2021a). Procyanidins concentrations of between 10.6 and 40.0 mg/g have been reported in sorghum, depending on the genotype. This is higher than that of blueberries, known for their high procyanidin content (Yu et al. 2018). Sorghum grain extracts rich in tannin polyphenolics were shown to have in vitro and in vivo inhibitory effects against α-amylase and α-glucosidase enzymes. This could decrease hyperglycemia in diabetics (Links et al. 2015).

2.2 Carotenoids

Carotenoids have many beneficial effects on human health. The most researched carotenoids in sorghum include lutein and zeaxanthin, which are the xanthophylls, which are also the major carotenoids in sorghum, and β-carotene. Varied concentrations of carotenoids have been reported in different studies. This may be due to a genotype effect and different extraction and detection methods. In one study, the total carotenoid content in a hundred sorghum genotypes varied from 2.12 to 85.46 μg/100 g. Nine genes were reported to be involved in carotenoid synthesis or degradation, which is a contributing factor to the variability of levels of carotenoids in the grains of sorghum (Cardoso et al. 2015).

2.3 Lipids and Vitamin E

Sorghum grain has low lipid content, though higher than most other cereal grains, and its lipid profile is similar to that of maize. The oil from sorghum grain have higher levels of oleic and stearic acids and lower levels of linoleic, myristic, and palmitoleic acids, making it less saturated than maize grain oil. The fact that oil from sorghum grain is high in unsaturated fatty acids may provide the health benefit of lipid-lowering properties (Khalid et al. 2022).

Sorghum grain fatty acids are considered as bioactive compounds with health benefits, especially the phytosterols and policosanols. These fatty acids occur in the lipid fraction of the grain as the lipid group of triacylglycerols with a high linoleic, oleic, and palmitic fatty acid content. Phytosterols are steroids that originate from plants, and β-sitosterol is the main phytosterol that was found in sorghum. Campesterol and stigmasterol were also isolated. The concentration of phytosterols is largely influenced by the environment, genotype, and extraction methods. Policosanols (a class of high-molecular-weight aliphatic alcohols) also have different types of bioactivity. Different policosanols (C28, C30, and C32) were isolated from sorghum, of which C28 policosanol was the predominant one (Wongwaiwech et al. 2020).

The vitamin E contents in sorghum were reported to vary significantly, between 280.7 and 2962.4 μg/100 g (wet basis). The α-, β-, γ-, and δ-tocopherols are the most studied in sorghum, with γ-tocopherol being the tocochromanol most frequently found, followed by α-tocopherol. The genotype and the growing environment both have a significant influence on the vitamin E content in sorghum grains (Cardoso et al. 2015).

2.4 Amines

There are two classes of amines, which are the biogenic amines and polyamines, and they represent a class of low-molecular-mass nitrogenous bases. The polyamines represent 60–100% of the total amines. Amines are considered to be bioactive compounds, and sorghum was reported to be a main source of polyamines. Spermine and spermidine are the most prevalent amines, followed by putrescine and cadaverine (Paiva et al. 2015).

2.5 Carbohydrates

The major carbohydrate in sorghum grain is starch, present in a granular form in the endosperm. Starch consists of two polysaccharides, amylose and amylopectin, and can be classified as waxy and non-waxy. Sorghum starch classified as non-waxy is composed of amylose (25%) and amylopectin (75%), while waxy sorghum consists almost entirely of amylopectin. Waxy sorghum is highly digestible by enzymes, while non-waxy starch exhibits resistance to enzymatic digestion. Proteins, tannins, and starch granules interact in the sorghum grains, forming complexes, leading to very poor starch digestibility. Sorghum has the poorest starch digestibility of all cereals. Sorghum genotypes with high phenolic and tannin content were found to be

associated with enzyme inhibition and starch molecule interaction. This impairs the starch digestibility, which increases the resistance of starch, which leads to a lower glycemic index of foods in which the starch is included (Moraes et al. 2018).

2.6 Fibers, Vitamins, and Minerals

Dietary fibers are mainly non-starch polysaccharides, which have numerous health benefits, such as positive effects on diabetes, tumors, and atherosclerosis. The whole grain of sorghum consists 10–25% of bran, of which 35–48% is insoluble dietary fiber (Miafo et al. 2019). Sorghum, being a fiber-rich food, has a low glycemic index, causing a slower and lower rise in blood glucose level (Moraes et al. 2018). Sorghum grains contain many and various types of vitamins and minerals. Minerals include Ca, Fe, K, Mg, P, and Zn (Motlhaodi et al. 2018). The most prevalent vitamins are of the B-complex such as pyridoxine (vitamin B6), riboflavin (vitamin B2), and thiamine (vitamin B1), and the fat-soluble vitamins such as A, D, E, and K (Przybylska et al. 2019).

2.7 Sorghum Protein

Cereal grains generally contain up to 20% proteins, but unlike animal protein, they have a low essential amino acid content. However, cereals are widely available and are a staple in many regions throughout the world, making them an attractive protein source and a source from which peptide nutraceuticals can be released, which could eliminate reactive oxygen (Szerszunowicz and Kłobukowski 2020). A wide range of protein content in sorghum grains have been reported (6–18%), of which the biggest percentage is storage protein, or prolamins. Sorghum prolamins are called kafirins and are located in the protein bodies of the endosperm. Sorghum kafirins make up a large portion of the total protein in whole kernels (48–70%) and an even higher portion (up to 80%) in decorticated kernels, while the rest consists of the albumins and globulins (Espinosa-Ramirez and Serna-Saldivar 2016. Discrete kafirin protein bodies are formed, and in mature endosperm, these bodies form a tight matrix with starch granules. These matrices affect the processing quality of sorghum and contribute to grain hardness and digestibility. Sorghum storage proteins are classified according to solubility, structure, amino acid composition, and molecular mass. They consist mainly of albumins (water extractable fraction), globulins (dilute salt extractable fraction), and kafirins or prolamins (alkali soluble fraction) (Wong et al. 2009).

Kafirins are hydrophobic proteins and are divided into three groups based on their molecular weight: α-kafirin (23–27 kDa), β-kafirin (16, 18, and 20 kDa), and γ-kafirin (28 kDa) (Espinosa-Ramirez and Serna-Saldivar 2016). The most dominant kafirin proteins are the α-kafirins, accounting for 70–80% of total storage proteins in sorghum grain (66–71% in opaque endosperm). They occur mostly in outer layers of grain and a reduction of α-kafirins is associated with a loss of vitreous endosperm texture (Wu et al. 2013). Sodium dodecyl sulfate polyacrylamide gel electrophoresis

(SDS-PAGE) has been used frequently in the past to separate the kafirins. With this technique, the α-kafirins resolve into two distinct bands with a molecular weight of 23 and 25 kDa, although these molecular weights can range from 22 to 27 kDa depending on different laboratory protocols and conditions (Cremer et al. 2014).

The second largest group of kafirins is the γ-kafirins. They make up a large part of the total kafirins in opaque endosperm (19–21%) and a smaller part (9–12%) in vitreous endosperm. They have a molecular weight of about 20 kDa. The β-kafirin comprises about 7–8% of total kafirins in vitreous endosperm (10–13% in opaque endosperm). The outside layers of the protein bodies contain most of the β- and γ-kafirins, where cross-links are formed between proteins via disulfide bonds (Oria et al. 2000). The very minor δ-kafirins (Mr = 15 kDa) form the fourth group, accounting for less than 1% of the mature grain total seed storage protein. Sorghum protein is poorly digestible, and the improvement of protein digestibility is still a major research goal in the sorghum fraternity. Protein digestibility is strongly related with the overall grain α- and γ-kafirin content (Wu et al. 2013; Elkonin et al. 2016). Dense, spherically shaped protein bodies are formed in the seed endosperm during grain development from the accumulation of the different kafirin groups. Kafirin protein bodies are quite compact, and on the periphery of these bodies, cross-links are formed. This probably causes the structural barriers preventing enzymatic access, which reduce the digestibility of these proteins (Oria et al. 2000).

Amino acids in grains are largely obtained from storage proteins, so this is a major factor determining the nutritional quality of the grain as a source of food for humans and animals. Both maize and sorghum proteins are deficient in the essential amino acids lysine and tryptophan, reducing the nutritional value of these cereals, which is an even bigger problem in sorghum than maize. Sorghum grain has low amount of essential amino acids like methionine, lysine, and isoleucine. The main amino acids in sorghum protein are histidine, leucine, phenylalanine, tyrosine, threonine, tryptophan, and valine (Mohapatra et al. 2019).

The γ-kafirins are rich in proline, cysteine, and histidine. The β-kafirins are rich in cysteine. As the δ-kafirin fractions are so small, it could be inconsequential in influencing sorghum grain quality traits. However, it is rich in the essential amino acid methionine, which is often deficient in cereal proteins, along with lysine, threonine, and tryptophan. If the δ-kafirin fraction can be increased, it could enhance the nutritional quality of sorghum proteins (Laidlaw et al. 2010).

2.8 Other Proteins in Sorghum

Various antifungal proteins have been identified in sorghum, including chitinase, glucanase, thionin, defensin, protease inhibitor, and ribosome-inactivating proteins. There are also several bioactive proteins in sorghum such as amylase and protease inhibitors, as well as glycine-rich RNA-binding proteins, protein kinases, and glutathione S-transferase isoenzymes. Proteins involved in lysine catabolism were also isolated from sorghum, which included lysine 2-oxoglutarate reductase and saccharopine dehydrogenase (Lin et al. 2013).

3 Genetic Improvement of Sorghum Nutraceutical Content

The nutraceutical content of sorghum can be improved in breeding programs if sufficient genetic variability is available to select for higher content. Natural genetic variation can be exploited, or genetic variability can be introduced from wild and close relatives, many of which are available from conserved genetic resources. Molecular techniques such as linkage and association mapping, and more recently, genome-wide association studies, can be used to understand the underlying genetic architecture of nutraceutical content. Sorghum mutant libraries are also proving very useful to use reverse genetics to determine traits that are useful for sorghum improvement, which could include nutraceutical content. Genetic engineering and genome editing techniques can now be applied to improve important traits, which already include some nutraceuticals.

3.1 Genetic Resources for Genetic Improvement of Nutraceuticals and Nutritional Value

The sorghum genus consists of *Sorghum bicolor.* This belongs to the subgenera of *Eu Sorghum*, within which there are three species: *Sorghum bicolor*, *Sorghum halepense* (Johnson grass), and *Sorghum propinquum*. The *S. bicolor* species consists of three subspecies: *bicolor*, *verticilliflorum*, and *drummondii* (Sudan grass). The *bicolor* subspecies has 5 races (bicolor, caudatum, durra, guinea, and kafir) and 10 intermediate races (Ananda et al. 2020). There are numerous global genetic resources of conserved sorghum worldwide and over 240,000 accessions are conserved in ex-situ gene banks. Most of these are cultivated accessions (98.3%) and a small percentage are wild weedy relatives (1.7%) (Upadhyaya et al. 2016), although there may be duplicates which yet have to be identified and managed. There are at least 20 sorghum gene banks in the world. The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) in India have a collection of 37,949 accessions from 92 countries (ICRISAT 2021). The United States Department of Agriculture (USDA) National Plant Germplasm System (NPGS) at the Plant Genetic Resources Conservation Unit in Griffin, GA, maintains a collection of over 45,000 accessions (USDA 2021). The Institute of Crop Science, Chinese Academy of Agricultural Sciences (ICS-CAAS), China, holds a collection of 18,263 accessions, and the National Bureau of Plant Genetic Resources (NBPGR) of India conserves 20,221 accessions (NBPGR 2021). Wild relatives are widely used in cereal breeding programs and likewise has potential in sorghum breeding to enhance yield potential and genetic diversity.

3.2 Subset Collections as Sources for Marker-Assisted Breeding

Significant research has been done on accessions collected from Ethiopia. Usually a high number of accessions is screened in different growing environments, and then a subset of this is genotyped using a genotyping by sequencing (GBS) approach. A

core subset is then identified which best represents the genetic diversity present in the whole germplasm collection. A subset of 374 accessions followed by another subset of 1425 Ethiopian landrace accessions were analyzed (Girma et al. 2019). After that, another subset of 387 Ethiopian germplasm accessions were comprehensively phenotyped and genomic characterization was done (Girma et al. 2020). The natural variation and genetic structure of these populations were determined. Single nucleotide sequence polymorphisms (SNP) associated with important traits can be used as molecular markers in breeding programs (Girma et al. 2019, 2020).

3.3 Sorghum Linkage and Association Mapping Resources

Natural variation in populations has been shaped by natural and artificial selection, and a combination of the two. Variation is probably linked to specific agronomic adaptive traits that have ecophysiological relevance. Ecogeographical and historical information also largely determine natural variation, and analysis of this variation can shed light on the evolutionary processes that led to the genetic variation patterns. Sorghum germplasm is known to have high levels of natural variation (Boyles et al. 2019) and is very suited for genetic analysis. In linkage mapping, recombination is generated by crossing parents, while in association mapping, historical recombination events are used. Linkage and association mapping are widely used to characterize the genetic architecture of traits. This includes information of how many loci are involved and how they are distributed, the gene action determining the traits, and linkage and allele frequency. With this knowledge, scientists can hypothesize on the genes that underlie variation in studied traits.

3.3.1 Linkage Mapping Resources

Many genetic traits in sorghum have been investigated through linkage mapping in the past 25 years, such as plant and flowering traits, pigmentation, drought and cold tolerance, and disease resistance (Xin et al. 2021). Almost no research has been reported on seed composition, including nutraceutical content, with the exception of kafirins.

3.3.2 Association Mapping Resources

There has been significant developments in the methods that can be used to generate high-density genome-wide markers. This has caused a shift towards genome-wide association studies (GWAS) from linkage mapping and gene association. The sorghum association panel is currently the most widely used sorghum GWAS resource. This panel was designed to capture and represent global plant diversity, and plant function and end uses. This association panel has genotyping by sequencing (GBS) marker data (Hu et al. 2019) and is available from the Germplasm Resources Information Network (GRIN). All the sorghum association panel accessions flower in temperate latitudes. This panel has been used for GWAS studies of many traits (Mural et al. 2020). There is also a bioenergy association panel, which is another global diversity panel that is available from GRIN, where the majority of the

accessions are tropical and photoperiod-sensitive. Although this panel consists mainly of tropical accessions, sweet and forage sorghums are also included. The bioenergy association panel has associated SNP data from GBS and whole-genome resequencing (Bellis et al. 2020; Lozano et al. 2021). A sorghum collection has been made available for phenotyping and association studies, consisting of 2000 georeferenced accessions, which were genotyped with GBS. Most of these accessions are also available from GRIN (Bellis et al. 2020). A large number of landrace accessions have been collected from Africa, for which GBS SNP data are available. There are entire GRIN collections from Niger, Nigeria, Senegal, Ethiopia, and Sudan, and for sweet sorghum.

3.3.3 Multi-parent Mapping Resources

Techniques such as linkage mapping are powerful and sensitive, while association mapping has the strength of power and sensitivity, which created an interest to combine the best of these techniques in different genetic approaches. This has led to the use of multi-parent mapping approaches, which include nested association mapping (NAM), backcross NAM, and multi-parent advanced generation intercross (MAGIC), all which have been applied in sorghum research (Boyles et al. 2019). The NAM approach is based on that which was developed for maize and now applied in many crops where multiple recombinant inbred line (RIL) families are developed that have diverse founders but share a common parent (Gage et al. 2020). RTx430 was selected as the common founder line for the global grain sorghum NAM resource (primarily because it has for decades been the most important public pollinator line), together with 10 diverse global founders (all part of the sorghum association panel), from which a total of more than 2200 RILs in 10 families were developed. The common germplasm between the association panel and the NAM allows the validation and comparison of the data from these two sources (Olatoye et al. 2020).

In the past, mainly biparental linkage families were used in research, but the aim of the MAGIC approach is to increase allelic diversity and at the same time to increase the power and specificity of quantitative trait loci (QTL) detection relative to GWAS. All founder lines contribute equally to the MAGIC in order to balance allele frequencies. One sorghum MAGIC resource has been developed and is currently available from the developers (Ongom and Ejeta 2018). There is also a backcross NAM (BCNAM) approach, where backcrosses are made and selection done to recover the common elite parent phenotype, which may have more of a breeding advantage (Jordan et al. 2011).

3.4 The Sorghum Mutant Library

Xin et al. (2008) developed a pedigreed mutant library for functional genomic studies by using a systematic approach in such a way that all mutations are captured and preserved. Individual seeds from an elite inbred line, BTx623 (used to generate the first sorghum reference genome), were treated with the chemical mutagen ethyl

methane sulfonate (EMS). Seeds were soaked in various concentrations of EMS to optimize the chances to obtain desirable mutations. This mutant library is a permanent resource, which can be used by scientists to test mutants under different growing conditions, to enhance breeding efforts. Currently, 6400 independent seed pools are available in this library to breeders and scientists. These are an excellent resource for sorghum breeding, as mutants with potentially useful traits can be selected for sorghum improvement programs (Xin et al. 2021). Purdue University has followed a similar approach to generate approximately 10,000 pedigreed seed pools (Addo-Quaye et al. 2018). A wide range of phenotypes that could be used for sorghum improvement have been identified and selected within this pedigreed mutant library. The traits and phenotypes identified in this mutant library have the potential to be used in sorghum breeding programs and in genomic studies. Both forward and reverse genetic resources have been developed that can be used to explore the mutant library traits (Wang et al. 2021).

Next-generation sequencing techniques have been developing rapidly, providing large numbers of DNA markers, which are also financially affordable. The output of high-quality DNA sequences has been increasing, and next-generation sequencing has become much cheaper. Whole genome sequencing of pedigreed mutants has become a resource for reverse genetics by searching gene mutations online. Next-generation sequencing has been used to annotate SNP markers. Mutant phenotypes of interest can be studied to map and identify causal mutations through mapping-by-sequencing or next-generation mapping (Hartwig et al. 2012).

Addo-Quaye et al. (2018) introduced mutations with EMS and then sequenced 586 mutants, where they identified 1,275,872 homozygous and 477,531 heterozygous mutations. In the sorghum genome, as is the case in other crops, the mutations often have deleterious effects. The sequenced data of the mutant collections can be a very useful resource for reverse genetics. These collections can be used together with other methods, such as genome-wide association and biparental mapping of QTLs of important agronomic traits, to validate candidate genes (Xin et al. 2021). Although very limited research has been done on the genetics of nutraceuticals so far (with the exception of kafirins) compared to the adaptive and yield-related traits, the resources are there, and they should be used to generate data for genetic improvement.

3.5 Genetic Engineering Approaches for Improving Nutritional Composition of Sorghum Grain

Recent years have been marked by significant progress in sorghum genetic transformation technologies that have made it feasible to use genetic engineering and genome editing methods to improve the nutritional properties of sorghum grain. Such progress is based on improved technologies of plant regeneration in sorghum tissue culture, in particular, the use of nutrient media containing increased concentrations of phosphate, proline, and asparagine, which reduces the release of phenolic pigments characteristic of cultivated sorghum tissues and increase embryogenic potential (Elkonin and Pakhomova 2000), as well as elevated levels of copper

ions, which improve the development of the root system of regenerants. Efficient media for plant regeneration allowed substantial improvement of genetic transformation methods of sorghum either through biolistic DNA delivery (Belide et al. 2017) or *Agrobacterium*-mediated genetic transformation (Do et al. 2016). In the latter case, significant improvements are due to the use of "hypervirulent" *A. tumefaciens* strain NTL_4 or another *Agrobacterium* strain containing specially designed Ti-plasmids with additional copies of *vir*-genes ("superbinary" vectors) or containing "helper" plasmids with additional *vir*-genes that enhance the transfer of T-DNA from *Agrobacterium* cells to sorghum cells. An important factor that also contributed to the increase in the efficiency of *Agrobacterium*-mediated genetic transformation in sorghum was the obtaining of an auxotrophic mutant LBA4404 Thy – incapable of growth on a media without thymidine, the use of which made it possible to significantly simplify the transformation procedure. The creation of binary vectors carrying the genes of morphogenetic regulators BABY BOOM and WUSCHEL, which promote the direct development of embryoids from scutellum cells of immature embryos, and thereby increase the number of regenerants and the frequency of transgenic plants, also had a significant effect (Che et al. 2022).

The main goals of genetic engineering approaches for improving nutritional value of sorghum grain are improvement of the digestibility of sorghum storage proteins, modification of starch content, enrichment of sorghum grain with essential amino acids and carotene, and decreasing the level of phytate that reduces bioavailability of minerals and phosphate. Some of these goals are being successfully realized currently, while the achievement of other goals is the task of future research.

3.5.1 Improvement of Kafirin Digestibility

Kafirins are resistant to proteolytic digestion, which is one of the main causes of reduced nutritional value of sorghum grain. This resistance to proteolytic digestion not only reduces their digestibility by animals and humans but also reduces the digestibility of grain starch, since undigested kafirins prevent the amylolytic cleavage of starch granules (Duressa et al. 2018).

Significant research has been devoted to the study of the factors that cause protein indigestibility. The reasons for kafirin resistance to protease digestion are multifactorial. It is assumed that kafirin resistance to protease digestion is caused by the chemical structure of kafirins, which are abundant with sulfur-containing amino acids capable to form S–S bonds that result in the formation of kafirin olygo- and polimers resistant to protease digestion; interactions of kafirins with non-kafirin proteins and nonprotein components such as polyphenols and polysaccharides; and the organization of different kafirins in the protein bodies, with the β- and γ-kafirin located at the periphery of the protein body, while the most abundant α-kafirin is located within the protein body. Such a spatial arrangement of kafirins in the protein bodies of sorghum, apparently, is a consequence of the prolamin synthesis pattern in the process of kernel development. In maize, it was found that starting from 4 to 5 days after fertilization, protein bodies accumulate γ- and β-prolamins, while later on, intensively accumulating α-prolamins moved γ- and β-prolamins to the periphery of the protein body. Due to the peripheral occurrence of γ-kafirin (which is

considered as the most stable to protease digestion) in the protein body, it is generally accepted that it reduces digestibility of α-kafirin, which comprises up to 80% of total endosperm kafirins (Duressa et al. 2018). In addition, γ-kafirin effectively forms oligo- or polymers of high molecular weight, which has high resistance to protease digestion.

This hypothesis was confirmed by the study of a mutant, P721Q, with a high level of lysine and improved digestibility of kafirins, obtained by chemical mutagenesis. In this mutant, the protein bodies had irregular shape, having deep invaginations, while normally they have a round shape. The γ-kafirin is located only in the bottom of these invaginations and does not form a continuous layer, which prevents digestion of α-kafirin in normal sorghum. It was suggested that such a structure of the protein bodies determines the high digestibility of proteins in the P721Q mutant and lines derived from it (Oria et al. 2000). Floury type endosperm grains are formed due to this mutation and lysine content is increased. It was therefore denoted with the symbol *hdhl* (*high digestibility high lysine*). Further investigation of this mutation using two-dimensional gel electrophoresis and mass spectrometry showed an increase of non-kafirin proteins (such as cytoskeleton and chaperone proteins, and the proteins involved in amino acids and carbohydrates synthesis) and a decrease in kafirin content in *hdhl* endosperm. The overexpression of chaperone proteins, which are probably involved in the repair of protein misfolding that was caused by the mutation, is in part responsible for increased lysine content of the P721-opaque sorghum (Benmoussa et al. 2015).

Studies have been undertaken on the chromosome localization of this mutation. In the research of Winn et al. (2009), with the analysis of a hybrid population obtained by crossing highly digestible line P850029 (derived from P721Q) and the wild type line Sureno, two QTLs were identified. These QTLs (both from the high-digestibility parent) were located on chromosome 1 in genomic regions within 20 cM from each other. They were in repulsion phase, meaning one QTL (locus 1 from the HD parent) unfavorably affects digestibility and one QTL (locus 2 from the HD parent) favorably affects digestibility. Protein digestibility can be increased if this linkage in repulsion is broken, which will allow the recombination of favorable alleles. It is noteworthy that another research group found that the proteinase inhibitor gene is located on chromosome 1 (Duressa et al. 2018). Perhaps QTLs identified by Winn et al. (2009) are linked to this gene.

In another investigation of chromosome localization of the *hdhl* mutation, one major QTL was identified on chromosome 5, in the 58 Mb region that overlaps with the genomic loci of the 17-kafirin gene cluster. In this research, the F_2 population generated from a cross between a normal line (BTx623) and a high-digestible mutant, P721Q was used (Massafaro et al. 2016). The F_2 plants with highly digestible protein displayed the unique protein body structure of P721Q. The results of this investigation confirms the data obtained by Wu et al. (2013) which showed that the *hdhl* phenotype in the P721Q mutant and lines derived from it is a consequence of a point mutation in one of the genes from the *k*1*C* family located on chromosome 5, encoding the 22 kDa α-kafirin. Sequencing revealed a point mutation in the nucleotide sequence that encodes the signal polypeptide responsible

for the packaging of α-kafirin inside the protein body (substitution at position 61 G→A) (Wu et al. 2013). This mutation resulted in the formation of a missense codon at the last amino acid of the signal sequence. Despite the fact that the mutation was found in only one of the genes of this cluster, its presence was sufficient to cause significant changes in the structure of protein bodies and the digestibility of kafirins. It was hypothesized that this mutation decreases the accumulation of α-kafirin in protein bodies that leads to a change in their ultrastructure and increases their sensitivity to the action of proteases (Wu et al. 2013).

In addition, a comparative analysis of the nucleotide sequences of all 27 kafirin genes in sorghum samples with high and low digestibility revealed four α-kafirin alleles localized on chromosome 5, which are closely associated with digestibility. Three alleles were associated with high digestibility (*Sobic.005G185600*, *Sobic.005G188800*, and *Sobic.005G189000*) and one with low digestibility (*Sobic.005G192801*). In silico predictive analysis showed the variants cause missense change in the amino acid sequences of the corresponding proteins (Duressa et al. 2020).

3.5.2 RNA Interference Technology

The use of biotechnology methods, in particular, RNA interference technology, opens up much broader prospects for obtaining mutants with improved digestibility of kafirins. This technology has been widely used for a long time to modify the synthesis of storage proteins as well as starch and other nutrients of endosperm in a number of cereals (Elkonin et al. 2016).

A number of research groups have been doing studies on the induction of RNA silencing of kafirin genes (Da Silva et al. 2011a, b; Kumar et al. 2012; Grootboom et al. 2014). RNA silencing was induced by genetic constructs carrying inverted repeats of several kafirin genes (α1, δ2, γ1, and γ2) separated by the *ADH*1 gene intron sequence. mRNA transcribed from this construct forms a double-stranded hairpin and underwent enzymatic degradation. The constructs were driven by the 19-kDa maize α-zein promoter (Da Silva et al. 2011a, b). In another study (Kumar et al. 2012), the genetic construct used to induce γ-kafirin silencing consisted of the complete *γ-KAFIRIN* gene sequence under the control of its own promoter. As a terminator, the sequence of the tobacco mosaic virus ribozyme gene was used. The expression of this gene should destroy γ-kafirin mRNA. In another construct, to induce α-kafirin silencing, α-kafirin inverted repeats separated by the intron sequence of the *Arabidopsis* gene encoding the D1 spliceosome protein were used; this construct was driven by the α-kafirin promoter (Kumar et al. 2012). Later, another complex construct consisting of inverted fragments of γ1-, γ2-, and δ-kafirin genes was used to obtain transgenic sorghum plants with silencing of kafirins (Grootboom et al. 2014).

The transgenic plants obtained in these experiments had a modified protein body structure which resembled that of the P721Q mutant, and an increased digestibility of kafirins, which was observed when both raw and cooked sorghum flour was treated with pepsin. One such example is the transgenic plants of Tx430, carrying a genetic construct for silencing α- and γ-kafirin. They were identified as having

improved in vitro protein digestibility (78% and 61% for the raw and cooked flour, respectively), while in the non-transgenic control, the digestibility varied in the ranges of 40–50% and 34–40%, respectively (Da Silva et al. 2011a). The genetic construct to suppress only δ- and γ-kafirins also improved the digestibility of raw flour but did not affect the digestibility of cooked flour. In the experiments of Kumar et al. (2012), the cooked flour from transgenic kernels carrying a genetic construct for γ-kafirin suppression did not differ from the non-transgenic control, while α-kafirin suppression improved the digestibility of such flour.

Based on the hypothesis that the surface location of γ-kafirin in protein bodies is the reason for the low digestibility of sorghum protein bodies, a genetic construct (NRKAFSIL) was created that should prevent the accumulation of the γ-kafirin without affecting the accumulation of other kafirins (Elkonin et al. 2016). This construct contained the fragment of the nucleotide sequence of γ-kafirin gene (GenBank Accession No: M73688) in direct and inverted orientation, separated by the maize ubiquitin intron sequence. The construct was driven by the 35S promoter. Using *Agrobacterium*-mediated genetic transformation, this construct was introduced into two sorghum cultivars: Zheltozernoe 10 (Zh10) and Avans. Electrophoretic spectra of endosperm proteins were compared before and after pepsin digestion, showing that in the transgenic plant, the amount of undigested kafirin monomers and total undigested protein was significantly lower than in the original non-transgenic line (by 1.7–1.9 times) when quantitative SDS-PAGE analysis was done. The level of digestibility reached 85–92%, while in the original line, this value was about 60%. The effect of increased digestibility of kafirins was traced to the T_4 generation; however, in some cases, it disappeared, possibly due to the instability of the introduced genetic construct or due to its silencing.

It should be noted that RNAi silencing of the γ-kafirin gene also caused a decrease in the content of α-kafirins. This phenomenon was especially pronounced in the Avans-1/18 transgenic line (Elkonin et al. 2021). At the same time, the functioning of the construct for the γ-kafirin gene silencing did not lead to a significant decrease of the total protein content of the grain compared to the original non-transgenic cultivar (14.3% vs. 15.5%). This fact, apparently, is a consequence of the balancing of protein synthesis in the kernels.

Another feature of transgenic sorghum lines with silencing of γ- and α-kafirins is the modification of the endosperm texture, in particular, the reduction of vitreous endosperm layer and formation of kernels containing floury endosperm characteristic to the P721Q mutant (Da Silva et al. 2011b; Kumar et al. 2012; Grootboom et al. 2014). Apparently, expression of RNAi genetic constructs also affects the formation of the protein-carbohydrate matrix, which caused formation of the vitreous endosperm layer. In experiments with the genetic construct for γ-kafirin gene silencing in variety Zh10, transgenic plants were obtained that had kernels with different endosperm types (Fig. 1).

Normal endosperm with a thick or thin vitreous layer, floury endosperm, or a modified type of endosperm in which the vitreous layer developed as sectors or spots surrounded by floury endosperm were seen (Elkonin et al. 2016). In these experiments, the kernels of some transgenic plants that exhibited a thick vitreous

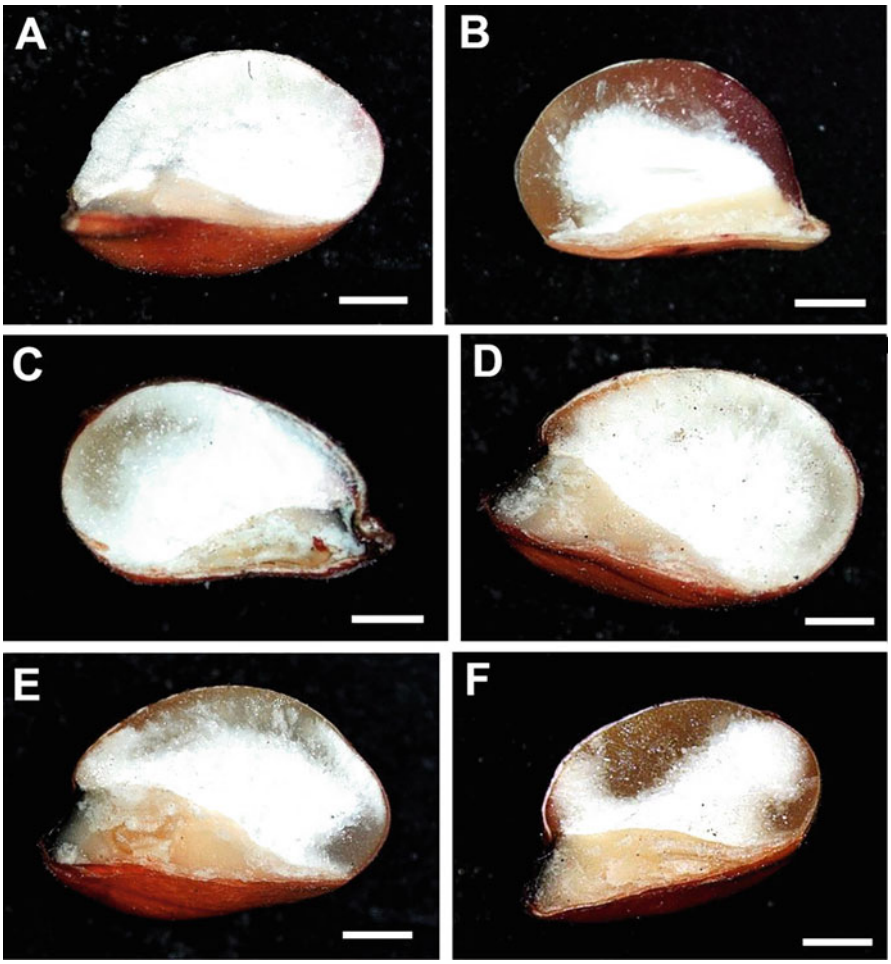


Fig. 1 Cross sections of kernels of transgenic sorghum plants with genetic construct for silencing of the γ-kafirin gene (Elkonin et al. 2016). (**a**) Kernel with floury endosperm set on the plant from T_3 generation; (**b**) kernel of original non-transgenic line Zh10 with thick vitreous endosperm (marked by arrows); (**c**–**e**) modified endosperm type with blurs and sectors of vitreous endosperm observed in kernels of different T_1 and T_2 plants; and (**f**–**h**) irregularly developed vitreous endosperm developed in kernels of T_2 and T_3 plants. Bar = 1 mm (© 2017 Elkonin LA, Italyanskaya JV, Panin VM, Selivanov NYu. Originally published in "Plant Engineering", InTech, Zagreb (Chroatia) under CC BY 3.0 license. Available from: https://doi.org/10.5772/intechopen.69973)

endosperm had a digestibility level as high as 92%, while the amount of undigested monomers was reduced by 17.5 times, and the amount of total undigested protein was 4.7 times less compared to the original lines. Previously, transgenic plants with similar endosperm texture were also observed in the Tx430 transgenic line with a construct for silencing α- and γ-kafirins.

It seems that the different types of endosperm that form are due to the particular expression of genetic constructs in the genome of the recipient line. In transgenic line Avans-1/18, in the T_1 generation, a revertant with vitreous endosperm and significantly reduced digestibility of kafirins was found (Elkonin et al. 2021). This revertant contained part of the NRKAFSIL genetic construct determining resistance to the selective agent (*bar* gene). At the same time, this revertant has a deletion in the *ubi*1 intron, which is a part of the construct for silencing of the *γ-KAFIRIN* gene, that, apparently, caused inefficiency of silencing and loss of characteristic features of mutation. These data indicate that the 588 bp *ubi*1 intron sequence can be used as a molecular marker in the screening of plants with high protein digestibility during hybridization of the Avans-1/18 mutant with various sorghum varieties.

The well-known correlation between the high digestibility of kafirins and floury endosperm was confirmed by these findings (Duressa et al. 2018). Such a correlation could hardly be explained by impaired synthesis of γ-prolamines (γ-kafirin), which are believed to play an important role in the interaction of protein bodies with starch granules, since mutants with impaired synthesis of α-kafirins also have a floury endosperm type (Da Silva et al. 2011b; Kumar et al. 2012; Grootboom et al. 2014). This correlation is possibly due to a violation of the formation of non-kafirin proteins in mutants with suppressed synthesis of α- or γ-kafirins, which form a protein matrix characteristic of vitreous endosperm. In mutants where kafirins are silenced, an increase in essential amino acid contents of lysine and threonine is seen, which appears to be due to increased synthesis of other proteins, including those with a higher content of essential amino acids. Thus, a significant increase of lysine content (an increase of 1.2 g/100 g protein compared to the non-transgenic control) was found in transgenic sorghum plants where α-, γ-, δ-kafirin genes were silenced due to the presence of complex genetic constructs for RNAi silencing and the presence of the lysine ketoglutarate reductase gene (which controls catabolism of free lysine) (Da Silva 2012).

In the transgenic plants of variety Zh10 containing the genetic construct for γ-kafirin silencing, the proportion of lysine increased 1.6–1.7 times (Elkonin et al. 2016); in the Avans-1/18 line, the increase was 75%, from 0.36% in the original line to 0.63%. It seems that the increase was caused by a decrease in amount of α-kafirins, which are poor in lysine and threonine, while the synthesis of other proteins was not affected. This caused an increase in the relative proportions of lysine and threonine. The suppression of γ-kafirin synthesis probably has no effect on the synthesis of proteins that are rich in lysine and threonine, but it prevents the accumulation of α-kafirins.

The use of RNA-interference technology, however, has a number of significant limitations: the genetic construct for silencing may undergo partial or complete destruction in the recipient genome (Elkonin et al. 2021), or its expression itself can undergo silencing. In addition, the functioning of the RNA silencing mechanism is an epigenetic process, and epigenetic processes in plants largely depend on environmental conditions (such as temperature, soil, and air humidity) (von Born et al. 2018). In the transgenic line Zh10, carrying the genetic construct for silencing

γ-kafirin, the level of digestibility of kafirins in plants grown in a field plot under natural moisture conditions was lower than in plants grown in outdoor vessels (Elkonin and Italyanskaya 2017). In some cases, off-target, nontarget, and unintended effects of genetic constructs for RNAi were observed (Christiaens et al. 2018). Besides these scientific problems, public concerns on using transgenic plants are a serious problem, which limits widespread use of sorghum lines carrying genetic constructs for RNA silencing for food and feed purposes.

3.5.3 Genome Editing Technologies

Site-directed mutagenesis using genetic constructs carrying the CRISPR/Cas system is one of the most effective technologies that are actively used to solve a variety of problems in genetics and plant breeding (Zhu et al. 2020). This approach allows the changing of the structure of the genes of plants, therefore, without introducing foreign genetic information, to change the plant metabolism in the necessary direction (Song et al. 2016). At the same time, in the offspring of mutants, due to recombination, it is possible to select plants that carry the induced mutation but are free from the genetic construct that induced it. As a result, the resulting mutants practically do not carry foreign genetic information; therefore, they are not transgenic organisms.

The CRISPR/Cas9 system includes Cas9 endonuclease and guide RNA (gRNA), which directs Cas9 endonuclease to the target nucleotide sequence (Song et al. 2016). The classical variant of the Cas9 nuclease recognizes the NGG-3′ Protospacer Adjacent Motif (PAM) sequence adjacent to the target (protospacer), which thus serves as the identification mark of the target in the edited genomic DNA. Cas9 endonuclease produces double-strand breaks in the target DNA located three nucleotides upstream of the PAM sequence. These breaks result in insertions or deletions at the target site, which can lead to frameshifts and null mutations.

The CRISPR-Cas system was successfully used to induce mutations in the nucleotide sequence of the 22 kDa signal polypeptide of α-kafirin (Li et al. 2018). Mutations were deletions ranging in size from 1 to 33 nucleotides, and, more rarely, insertions ranging in size from 1 to 16 nucleotides. In the kernels of T_1 and T_2 plants, a reduced level of α-kafirin and an altered structure of protein bodies were observed; some T_2 plants had higher protein digestibility and increased lysine levels. These results indicate the promise of using genome editing techniques to improve the nutritional value of sorghum grain.

Edited sorghum plants with mutations in the nucleotide sequences of the β- and γ-kafirin gene have also been reported (Massel et al. 2022). In this study, high efficiency of endogenous *U6* promoters (in particular, *SbU62.3* promoter) was shown to improve gene editing efficiency in sorghum of up to 90% of experimental plants.

A series of binary vectors (pC1-pC4) were created for site-directed mutagenesis of genes encoding α- and γ-kafirins (Gerashchenkov et al. 2021). These vectors contain the *Cas9* endonuclease gene under the control of the *ubi*1 promoter and gRNA nucleotide sequences that are complementary to target sites encoding signal polypeptides of α- and γ-kafirins (the *k*1*C*5 and *gKAF*1 genes, respectively). Using *Agrobacterium*-mediated genetic transformation, the constructs for site-directed mutagenesis of these genes were introduced into the genome of cv. Avans. Using

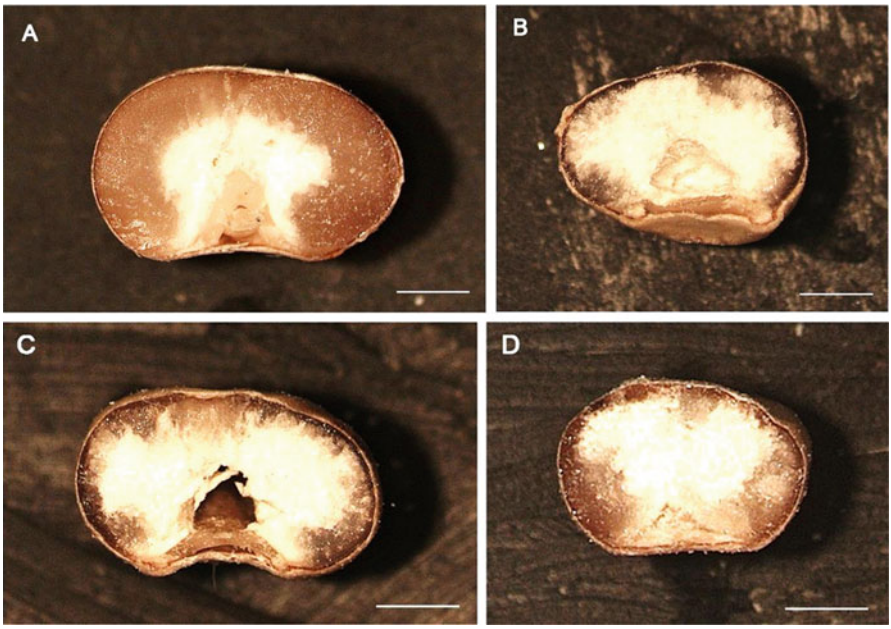


Fig. 2 Cross sections of kernels set on the panicle of the sorghum plant carrying a genetic construct for *α-KAFIRIN (k*1*C*5*)* gene editing target ((**b**), (**c**), (**d**)); ((**a**) – kernel of original cv. Avans). Bar 1 mm (Amer J Plant Sci, 2021, 12: 1276–1287, with permission)

the pC2 vector (to induce mutations in the *k*1*C*5 gene), regenerants (T_0) with kernels with modified endosperm texture were obtained, in which a significant reduction in the vitreous endosperm was observed (Fig. 2).

This demonstrates the efficiency of genome editing approaches for improvement of nutritional value of sorghum grain. The successful results obtained in these studies demonstrate the validity of the choice of kafirin signaling polypeptides as targets for improving the nutritional value of sorghum grain. However, it is likely that this approach is only one of the ways to use the CRISPR/Cas method to improve the nutritional value of sorghum grain, and other approaches will be used to solve this problem in future experiments.

3.5.4 Synthetic Biology Approaches

With the help of synthetic biology approaches, transgenic lines with increased protein digestibility and increased protein content were obtained that contain an artificially synthesized β-kafirin gene (Liu et al. 2019). This gene encoded modified β-kafirin protein with additional proteolytic cleavage sites that should improve its digestibility. Some of the resulting transgenic lines had higher protein content in the seeds (by 11–37%) and a higher digestibility of kafirins (by 11–21%) compared to the non-transgenic original variety. The protein bodies had an irregular shape with invaginations characteristic for highly digestible sorghum lines.

In order to increase the content of lysine in the grain, transgenic sorghum lines carrying the gene encoding the high-lysine protein of barley, gordotinin, were obtained. *Agrobacterium*-mediated genetic transformation was used to introduce the gene encoding the high lysine analog (HTl2 protein) of the *Hordeum vulgare* α-hordothionin protein under the control of the 27 kDa maize γ-zein promoter and terminator into the genome of two sorghum lines, P898012 and PHI391. The *A. tumefaciens* strain LBA4404 contained a "super-binary" vector with two unlinked T-DNA cassettes. The one cassette contained the lysine-rich *HTl2* gene with the second cassette containing a herbicide-resistant *bar* gene as a selectable marker. The two different T-DNA cassettes in the co-transformation vector allows the segregation of the marker and trait genes in the progeny of the primary transformants. This also allows the elimination of the marker gene and consequently marker-free transgenic plants are obtained. Three high levels of the HTl2 protein were expressed in the grain of the five independent transgenic events that were co-transformed with both genes, with a 40–60% increase of lysine.

Provitamin A is very important to human health, but sorghum grain has low levels of β-carotene or provitamin A. In order to increase β-carotene content in sorghum grain, the genetic construct encoding a number of enzymes involved in the carotenoid biosynthesis pathway has been introduced into the genome of sorghum line Tx430 (Che et al. 2016). These enzymes are: 1-deoxyxylulose 5-phosphate synthase, the precursor for carotenoid biosynthesis, *Zea mays* phytoene synthase 1, and the *Pantoea ananatis* carotene desaturase, involved in synthesis of phytoene and lycopene, respectively, which are the β-carotene precursors. This introduction resulted in an increase ofβ-carotene levels in mature seeds of transgenic plants of up to 9.1 μg/g (compared to 0.5 μg/g in non-transgenic control seeds). The β-carotene in plants can degrade during storage due to oxidation. To counter this effect, the barley *HGGT* gene encoding homogentisate geranylgeranyl transferase was introduced into the same genetic construct. Homogentisate geranylgeranyl transferase is involved in synthesis of vitamin E, and this vitamin has strong antioxidant effects. It was found that co-expression of the *HGGT* gene stacked with carotenoid biosynthesis genes, enhanced all-trans-β-carotene accumulation, and reduced β-carotene oxidative degradation. This led to stable provitamin A levels in sorghum seeds. Field trials of transgenic plants with increased carotene content showed that the all-trans-β--carotene levels were increased by about 20-fold in transgenic lines compared to the non-transgenic controls (Che et al. 2019). This is an example of how effective genetic engineering can be for modifying plant metabolism to meet human needs.

Thus, development of genetic transformation techniques allowed the development of sorghum lines with significantly improved protein digestibility, increased content of lysine and other essential amino acids, and β-carotene. However, there still remains a number of unsolved problems. Future research should focus on the development of lines with reduced phytate content, which lowers bioavailability of Fe, Zn, and phosphate, as well as lines which have high protein digestibility of seed storage proteins, and vitreous endosperm. Starch digestibility is also an important question because improving the caloric value of staple food is of great importance. Identification of the naturally occurring allele of pullulanase (*SbPUL-RA*) – starch

debranching enzyme – that confers significantly higher in vitro starch digestibility (Gilding et al. 2013) shows that this trait may be improved using a genome editing approach. These innovations, coupled with development of genomics of nutritive traits, will significantly improve the gene pool of existing varieties and hybrids of sorghum and make this important crop more in demand in world agricultural production.

4 Conclusion

Sorghum is an underutilized cereal, well adapted to changing climatic conditions, which are causing increased temperatures and an increased frequency of drought spells. The inclusion of especially whole grain sorghum in the diet may help avoid chronic lifestyle illnesses. The inclusion of sorghum grain as a regular part of the human diet has the potential to reduce the risk of cardiovascular diseases, some types of cancer, and type II diabetes. Sorghum has highly resistant starch, a high fiber content, high levels of bioactive compounds, and kafirin protein with potential benefits. The bioactivity of sorghum grain is influenced by phenolic compounds such as phenolic acid, flavonoids, stilbenes, and tannins, and it is rich in procyanidins (condensed tannins) and 3-deoxyantocyanidins. Sorghum tannins, also called proanthocyanidins, have shown anti-inflammation and anticancer properties. Sorghum also contains B-complex vitamins, fat-soluble A, D, E, and K vitamins, and minerals such as potassium, phosphorus, magnesium, and zinc. The generally high antioxidant activity of sorghum grain could contribute towards combating diseases associated with oxidative stress. Sorghum grain also has cholesterol-lowering and antimicrobial properties. It also was shown to improve glucose metabolism, which could positively affect diabetes. Sorghum is gluten-free with a high fiber content, which could benefit celiac disease patients. Sorghum could therefore be included in functional foods as a source of bioactive ingredients. Of the bioactive compounds in sorghum, phenolic compounds have been the main focus for research, so there is still a lot that is not known about the other compounds. With developing technology, new bioactive compounds may also be detected.

All the studies reported so far found significant natural genetic variation for measured bioactive compounds, although the growing environment also consistently had a large influence on the expression of compounds. As was reported above, there are many untapped genetic resources available in the form of germplasm collections, which can be screened for naturally high occurrence of bioactive compounds. Significant research has been done on the genetics of kafirin protein indigestibility and to a lesser extent on compounds like carotenoids. There is, however, a large gap in the knowledge of the genetics underlying the expression of other bioactive compounds in sorghum. Linkage and association mapping, GWAS and mutant libraries have been used to research the genetic architecture of especially adaptive traits, and yield and related traits, and genetic engineering has been applied to contribute to improved protein indigestibility and carotene levels in sorghum. There is, however, scant data on the genetics of seed composition, bioactive compounds, and traits linked to the

nutritional value of sorghum grain in general. When more is known about the genes linked to nutraceuticals and the gene action involved in expression of nutraceuticals, conventional breeding can be used to select and cross parents with good nutraceutical content, markers can be developed for marker-assisted selection, and genomics and genetic engineering could be used to improve nutraceutical content. Sorghum research is lagging behind that of other commercial crops such as maize, rice, and wheat. Likewise, research on the nutritional value of grains is lagging behind research on yield and yield-related traits, adaptive traits, and disease resistance. Databases should be generated to integrate all available information on seed nutraceutical and other nutritional components, which can be used by the sorghum community for sorghum improvement and genetic studies.

References

Addo-Quaye C, Tuinstra M, Carraro N, Weil C, Dilkes BP (2018) Whole-genome sequence accuracy is improved by replication in a population of mutagenized sorghum. G3 8:1079–1094. https://doi.org/10.1534/g3.117.300301

Ananda GKS, Myrans H, Norton SL, Gleadow R, Furtado A, Henry RJ (2020) Wild sorghum as a promising resource for crop improvement. Front Plant Sci 11:1108. https://doi.org/10.3389/fpls.2020.01108

Belide S, Vanhercke T, Petrie JR, Singh SP (2017) Robust genetic transformation of sorghum (*Sorghum bicolor* L.) using differentiating embryogenic callus induced from immature embryos. Plant Methods 13:109. https://doi.org/10.1186/s13007-017-0260-9

Bellis ES, Kelly EA, Lorts CM, Gao H, DeLeo VL, Rouhan G, Budden A, Bhaskara GB, Hu Z, Muscarella R, Timko MP, Nebie B, Runo SM, Chilcoat ND, Juenger TE, Morris GP, dePamphilis CW, Lasky JR (2020) Genomics of sorghum local adaptation to a parasitic plant. Proc Natl Acad Sci U S A 117:4243–4251. https://doi.org/10.1073/pnas.1908707117

Benmoussa M, Chandrashekar A, Ejeta G, Hamaker BR (2015) Cellular response to the high protein digestibility/high-lysine (*hdhl*) sorghum mutation. Planta 241:70–77. https://doi.org/10.1016/j.plantsci.2015.08.025

Bouargalne Y, Ben Mrid R, Bouchmaa N, Zouaoui Z, Benmrid B, Kchikich A, El Omari R, Kabach I, Mohamed N (2022) Genetic diversity for agromorphological traits, phytochemical profile, and antioxidant activity in Moroccan sorghum ecotypes. Sci Rep 12:5895. https://doi.org/10.1038/s41598-022-09810-9

Boyles RE, Brenton ZW, Kresovich S (2019) Genetic and genomic resources of sorghum to connect genotype with phenotype in contrasting environments. Plant J 97:19–39. https://doi.org/10.1111/tpj.14113

Bröhan M, Jerkovic V, Collin S (2011) Potentiality of red sorghum for producing stilbenoid-enriched beers with high antioxidant activity. J Agric Food Chem 59:4088–4094. https://doi.org/10.1021/jf1047755

Burdette A, Garner PL, Mayer EP, Hargrove JL, Hartle DK, Greenspan P (2010) Anti-inflammatory activity of select sorghum (*Sorghum bicolor*). Brans J Med Food 13:879–887. https://doi.org/10.1089/jmf.2009.0147

Cabrera AH, Luzardo I, Ramirez AK, Morales E, Campos R, Gaytan (2020) Effect of the nixtamalization process on the protein bioaccessibility of white and red sorghum flours during in vitro gastrointestinal digestion. Food Res Int 134:1–13. https://doi.org/10.1016/j.foodres.2020.109234

Cardoso L d M, Pinheiro SS, da Silva LL, de Menezes CB, de Carvalho CW, Tardin FD, Queiroz VA, Martino HS, Pinheiro-Sant'Ana HM (2015) Tocochromanols and carotenoids in sorghum

(*Sorghum bicolor* L.): diversity and stability to the heat treatment. Food Chem 172:900–908. https://doi.org/10.1016/j.foodchem.2014.09.117

Chaudhari SP, Powar PV, Pratapwar MN (2017) Nutraceuticals: a review. World J Pharm Sci 6:681–739. https://doi.org/10.1007/s13555-018-0221-x

Che P, Zhao Z-Y, Glassman K, Dolde D, Hu TX, Jones TJ, Gruis DF, Obukosia S, Wambugu F, Albertson MC (2016) Elevated vitamin E content improves all-trans beta-carotene accumulation and stability in biofortified sorghum. Proc Natl Acad Sci U S A 113:11040–11045. https://doi.org/10.1073/pnas.1605689113

Che P, Zhao Z-Y, Hinds M, Rinehart K, Glassman K, Albertsen M (2019) Evaluation of agronomic performance of β-carotene elevated sorghum in confined field conditions. In: Zhao Z-Y, Dahlberg J (eds) Sorghum: methods and protocols, Methods in molecular biology, vol 1931. Springer Science+Business Media, LLC, pp 209–220

Che P, Wu E, Simon MK, Anand A, Lowe L, Yang M, Albertsen MC, Gordon-Kamm W, Jones TJ (2022) *Wuschel*2 enables highly efficient CRISPR/Cas-targeted genome editing during rapid de novo shoot regeneration in sorghum. Commun Biol 5:344. https://doi.org/10.1038/s42003-022-03308-w

Choi SC, Kim JM, Lee YG, Kim C (2019) Antioxidant activity and contents of total phenolic compounds and anthocyanins according to grain colour in several varieties of *Sorghum bicolor* (L.) Moench. Cereal Res Commun 47:228–238. https://doi.org/10.1556/0806.47.2019.014

Christiaens O, Dzhambazova T, Kostov K, Arpaia S, Reddy Joga M, Urru I, Sweet J, Smagghe G (2018) Literature review of baseline information on RNAi to support the environmental risk assessment of RNAi-based GM plants. EFSA Supporting Publication 2018:EN-1424. https://doi.org/10.2903/sp.efsa.2018.EN-1424

Cox S, Noronha L, Herald T, Bean S, Lee SH, Perumal R, Wang W, Smolensky D (2019) Evaluation of ethanol-based extraction conditions of sorghum bran bioactive compounds with downstream anti-proliferative properties in human cancer cells. Heliyon 5:e015891–e015897. https://doi.org/10.1016/j.heliyon.2019.e01589

Cremer JE, Godwin ID, Liu L, Wang D, Bean SR, Tilley M, Wilson JD, Kaufman RC, Vu TH, Gilding EK, Godwin ID, Wang D (2014) Impacts of kafirin allelic diversity, starch content, and protein digestibility on ethanol conversion efficiency in grain sorghum. Cereal Chem 91:218–227. https://doi.org/10.1094/cchem-04-13-0068-r

Da Silva LS, Taylor J, Taylor JR (2011a) Transgenic sorghum with altered kafirin synthesis: kafirin solubility, polymerization, and protein digestion. J Agric Food Chem 59:9265–9270. https://doi.org/10.1021/jf201878p

Da Silva LS, Jung R, Zhao ZY, Glassman K, Grootboom AW, Taylor J, Taylor JRN (2011b) Effect of suppressing the synthesis of different kafirin sub-classes on grain endosperm texture, protein body structure and protein nutritional quality in improved sorghum lines. J Cereal Sci 54:160–167. https://doi.org/10.1016/j.jcs.2011.04.009

De Oliveira KG, Vieira VA, Lanamar C, de Morais L, Pinheiro HM, Anunciacao PC, Beserra C, da Silva EC, Barros F (2017) Effect of the storage time and temperature on phenolic compounds of sorghum grain and flour. Food Chem 216:390–398. https://doi.org/10.1016/j.foodchem.2016.08.047

Da Silva LS (2012) Transgenic sorghum: Effects of altered kafirin synthesis on kafirin polymerisation, protein quality, protein body structure and endosperm texture. [thesis]. Pretoria, South Africa: Department of Food Science, Faculty of Natural and Agricultural Sciences, University of Pretoria, South Africa

Do PT, Lee H, Mookkan M, Folk WR, Zhang ZJ (2016) Rapid and efficient *Agrobacterium*-mediated transformation of sorghum (*Sorghum bicolor*) employing standard binary vectors and bar gene as a selectable marker. Plant Cell Rep 35:2065–2076. https://doi.org/10.1007/s00299-016-2019-6

Duressa D, Weerasoriya D, Bean SR, Tilley M, Tesso T (2018) Genetic basis of protein digestibility in grain sorghum. Crop Sci 58:2183–2199. https://doi.org/10.2135/cropsci2018.01.0038

Duressa D, Bean S, Amand PS, Tesso T (2020) Identification of variant α-kafirin alleles associated with protein digestibility in grain sorghum. Crop Sci 60:2467–2478. https://doi.org/10.1002/csc2.20198

Elkonin LA, Italyanskaya YV (2017) *In vitro* digestibility of storage endosperm proteins of transgenic sorghum plants carrying genetic construct for silencing of the gamma-kafirin gene. Adv Cur Nat Sci 12:96–100. [In Russian]

Elkonin LA, Pakhomova NV (2000) Influence of nitrogen and phosphorus on induction embryogenic callus of sorghum. Plant Cell Tissue Organ Cult 61:115–123. https://doi.org/10.1023/A:1006472418218

Elkonin LA, Italianskaya JV, Domanina IV, Selivanov NY, Rakitin AL, Ravin NV (2016) Transgenic sorghum with improved digestibility of storage proteins obtained by *Agrobacterium*-mediated transformation. Russ J Plant Physiol 63:678–689

Elkonin LA, Panin VM, Kenzhegulov OA, Sarsenova SK (2021) RNAi-mutants of *Sorghum bicolor* (L.) Moench with improved digestibility of seed storage proteins. In: Jimenez-Lopez JC (ed) Grain and seed proteins functionality. IntechOpen Ltd., London, UK. https://doi.org/10.5772/intechopen.96204

Espinosa-Ramirez J, Serna-Saldivar SO (2016) Functionality and characterization of kafirin-rich protein extracts from different whole and decorticated sorghum genotypes. J Cereal Sci 70:57–65. https://doi.org/10.1016/j.jcs.2016.05.023

Espitia-Hernández P, Chávez González ML, Ascacio-Valdés JA, Dávila-Medina D, Flores-Naveda A, Silva T, Chacón XR, Sepúlveda L (2022) Sorghum (*Sorghum bicolor* L.) as a potential source of bioactive substances and their biological properties. Crit Rev Food Sci Nutr 62:2269–2280. https://doi.org/10.1080/10408398.2020.1852389

FAOSTAT (2019) Food and agriculture data. Food and Agriculture Organization of the United Nations. http://www.fao.org/faostat/en

Gage JL, Monier B, Giri A, Buckler ES (2020) Ten years of the maize nested association mapping population: impact, limitations, and future directions. Plant Cell 32:2083–2093. https://doi.org/10.1105/tpc.19.00951

Gerashchenkov G, Elkonin L, Gerashchenkov K, Rozhnova N, Hiekel S, Kumlehn J, Chemeris AV (2021) Binary vector construction for site-directed mutagenesis of kafirin genes in sorghum. Am J Plant Sci 12:1276–1287. https://www.scirp.org/journal/ajps

Gharaati S (2019) Extraction techniques of phenolic compounds from plants. In: Soto-Hernandez-M, Garcia-Mateos R, Palma-Tenango M (eds) Plant physiological aspects of phenolic compounds. IntechOpen. https://doi.org/10.5772/intechopen.84705

Gilding EK, Frère CH, Cruickshank A, Rada AK, Prentis PJ, Mudge AM, Mace ES, Jordan DR, Godwin ID (2013) Allelic variation at a single gene increases food value in a drought-tolerant staple cereal. Nat Commun 4:1483. https://doi.org/10.1038/ncomms2450

Girma G, Nida H, Seyoum A, Mekonen M, Nega A, Lule D, Dessalegn K, Bekele A, Gebreyohannes A, Adeyanju A, Tirfessa A, Ayana G, Taddese T, Mekbib F, Belete K, Tesso T, Ejeta G, Mengiste T (2019) A large-scale genome-wide association analyses of Ethiopian sorghum landrace collection reveal loci associated with important traits. Front Plant Sci 10:691. https://doi.org/10.3389/fpls.2019.00691

Girma G, Nida H, Tirfessa A, Lule D, Bejiga T, Seyoum A, Mekonen M, Nega A, Dessalegn K, Birhanu C, Bekele A, Gebreyohannes A, Ayana G, Tesso T, Ejeta G, Mengiste T (2020) A comprehensive phenotypic and genomic characterization of Ethiopian sorghum germplasm defines core collection and reveals rich genetic potential in adaptive traits. Plant Genome 13:e20055. https://doi.org/10.1002/tpg2.20055

Grootboom AW, Mkhonza NL, Mbambo Z, O'Kennedy MM, da Silva LS, Taylor J, Taylor JRN, Chikwamba R, Mehlo L (2014) Co-suppression of synthesis of major α-kafirin sub-class together with γ-kafirin-1 and γ-kafirin-2 required for substantially improved protein digestibility in transgenic sorghum. Plant Cell Rep 33:521–537. https://doi.org/10.1007/s00299-013-1556-5

Hartwig B, James GV, Konrad K, Schneeberger K, Turck F (2012) Fast isogenic mapping-by-sequencing of ethyl methanesulfonate induced mutant bulks. Plant Physiol 160:591–600. https://doi.org/10.1104/pp.112.200311

Hu Z, Olatoye MO, Marla S, Morris GP (2019) An integrated genotyping-by-sequencing polymorphism map for over 10,000 sorghum genotypes. Plant Genome 12:18044. https://doi.org/10.3835/plantgenome2018.06.0044

Hu J, Zhang L, Lin W, Tang W, Chan FKL, Ng SC (2021) Review article: probiotics, prebiotics and dietary approaches during COVID-19 pandemic. Trends Food Sci Technol 108:187–196. https://doi.org/10.1016/j.tifs.2020.12.009

ICRISAT (2021) Genetic resources. International Crops Research Institute for the Semi-Arid Tropics. http://exploreit.icrisat.org/profile/genetic%20resources/67

Iyabo OO, Ibiyinka O, Abimbola O (2018) Comparative study of nutritional, functional and anti-nutritional properties of white *Sorghum bicolor* (sorghum) and *Pennisetum glaucum* (pearl millet). Int J Eng Techn Manag Res 5:151–158. https://doi.org/10.29121/ijetmr.v5.i3.2018.187

Jiang Y, Zhang H, Qi X, Wu G (2020) Structural characterization and antioxidant activity of condensed tannins fractionated from sorghum grain. J Cereal Sci 92:102918. https://doi.org/10.1016/j.jcs.2020.102918

Jordan DR, Mace ES, Cruickshank AW, Hunt CH, Henzell RG (2011) Exploring and exploiting genetic variation from unadapted sorghum germplasm in a breeding program. Crop Sci 51:1444–1457. https://doi.org/10.2135/cropsci2010.06.0326

Kadri F, Messaoud B, Noureddine B, Bernard W (2017) Free radical scavenging activity correlated with phenolic contents in aqueous methanolic extracts of three Algerian cultivars of sorghum: *Sorghum bicolor* (L.) Moench. ElWahat J Res Stud 10:166–180. https://doi.org/10.54246/1548-010-002-007

Khalid W, Ali A, Arshad MS, Afzal F, Akram R, Siddeeg A, Kousar S, Rahim MA, Aziz A, Maqbool Z, Saeed S (2022) Nutrients and bioactive compounds of *Sorghum bicolor* L. used to prepare functional foods: a review on the efficacy against different chronic disorders. Int J Food Prop 25:1045–1062. https://doi.org/10.1080/10942912.2022.2071293

Kumar T, Dweikat I, Sato S, Ge Z, Nersesian N, Chen H, Elthon T, Bean S, Ioerger BP, Tilley M, Clemente T (2012) Modulation of kernel storage proteins in grain sorghum (*Sorghum bicolor* (L.) Moench). Plant Biotechnol J 10:533–544. https://doi.org/10.1111/j.1467-7652.2012.00685.x

Kumari PK, Umakanth AV, Narsaiah TB, Uma A (2021) Exploring anthocyanins, antioxidant capacity and a-glucosidase inhibition in bran and flour extracts of selected sorghum genotypes. Food Biosci 41:100979. https://doi.org/10.1016/j.fbio.2021.100979

Laidlaw H, Mace E, Williams S, Sakrewski K, Mudge A, Prentis P, Jordan D, Godwin I (2010) Allelic variation of the β-, γ- and δ-kafirin genes in diverse sorghum genotypes. Theor Appl Genet 121:1227–1237. https://doi.org/10.1007/s00122-010-1383-9

Li A, Jia S, Yobi A, Ge Z, Sato SJ, Zhang C, Angelovic R, Clemente TE, Holding DR (2018) Editing of an alpha-kafirin gene family increases digestibility and protein quality in sorghum. Plant Physiol 177:1425–1438. https://doi.org/10.1104/pp.18.00200

Li Z, Zhao X, Zhang X, Liu H (2021a) Bioactive compounds and biological activities of sorghum grains. Foods 10:2868. https://doi.org/10.3390/foods10112868

Li M, Xu T, Zheng W, Gao B, Zhu H, Xu R, Deng H, Wang B, Wu Y, Sun X, Zhang Y, Yu LL (2021b) Triacylglycerols compositions, soluble and bound phenolics of red sorghums, and their radical scavenging and anti-inflammatory activities. Food Chem 340:128123. https://doi.org/10.1016/j.foodchem.2020.128123

Lin P, Ho J, Bun T, Sai V, Ho M, Xia L (2013) A sorghum xylanase inhibitor-like protein with highly potent antifungal, antitumor and HIV-1 reverse transcriptase inhibitory activities. Food Chem 141:2916–2922. https://doi.org/10.1016/j.foodchem.2013.04.013

Links MR, Taylor J, Kruger MC, Taylor JRN (2015) Sorghum condensed tannins encapsulated in kafirin microparticles as a nutraceutical for inhibition of amylases during digestion to attenuate hyperglycaemia. J Funct Foods 12:55–63. https://doi.org/10.1016/j.jff.2014.11.003

Liu G, Gilding EK, Kerr ED, Schulz BL, Tabet B, Hamaker BR, Godwin ID (2019) Increasing protein content and digestibility in sorghum grain with a synthetic biology approach. J Cereal Sci 85:27–34. https://doi.org/10.1016/j.jcs.2018.11.001

Lozano R, Gazave E, Dos Santos JPR, Stetter MG, Valluru R, Bandillo N, Fernandes SB, Brown PJ, Shakoor N, Mockler TC, Cooper EA, Taylor Perkins M, Buckler ES, Ross-Ibarra J, Gore MA (2021) Comparative evolutionary genetics of deleterious load in sorghum and maize. Nat Plants 7:17–24. https://doi.org/10.1038/s41477-020-00834-5

Luo M, Hou F, Dong L, Huang F, Zhang R, Su D (2020) Comparison of microwave and high-pressure processing on bound phenolic composition and antioxidant activities of sorghum hull. Int J Food Sci Technol 55:3190–3202. https://doi.org/10.1111/ijfs.14583

Massafaro M, Thompson A, Tuinstra M, Dilkes B, Weil CF (2016) Mapping the increased protein digestibility trait in the high-lysine sorghum mutant P721Q. Crop Sci 56:2647–2651. https://doi.org/10.2135/cropsci2016.03.0188

Massel K, Lam Y, Hintzsche J, Lester N, Botella JR, Godwin ID (2022) Endogenous U6 promoters improve CRISPR/Cas9 editing efficiencies in *Sorghum bicolor* and show potential for applications in other cereals. Plant Cell Rep 41:489–492. https://doi.org/10.1007/s00299-021-02816-z

Miafo APT, Koubala BB, Kansci G, Muralikrishna G (2019) Free sugars and non-starch polysaccharides–phenolic acid complexes from bran, spent grain and sorghum seeds. J Cereal Sci 87:124–131. https://doi.org/10.1016/j.jcs.2019.02.002

Mohapatra D, Patel AS, Kar A, Deshpande SS, Kumar M (2019) Effect of different processing conditions on proximate composition,anti-oxidants, anti-nutrients and amino acid profile of grain sorghum. Food Chem 271:129–135. https://doi.org/10.1016/j.foodchem.2018.07.196

Moraes E, da Silva R, Alves S, Vieira VA, Ludemann R, Borck PC, Magalhaes E, Marostica MR (2018) Whole sorghum flour improves glucose tolerance, insulin resistance and preserved pancreatic islets function in obesity diet-induced rats. J Funct Foods 45:530–540. https://doi.org/10.1016/j.jff.2017.03.047

Motlhaodi T, Bryngelsson T, Chite S, Fatih M, Ortiz R, Geleta M (2018) Nutritional variation in sorghum [*Sorghum bicolor* (L.) Moench] accessions from southern Africa revealed by protein and mineral composition. J Cereal Sci 83:123–129. https://doi.org/10.1016/j.jcs.2018.08.010

Mural RV, Grzybowski M, Miao C, Damke A, Sapkota S, Boyles RE, Fernandez MGS, Schnable PS, Sigmon B, Kresovich S, Schnable JC (2020) Meta-analysis identifies pleiotropic loci controlling phenotypic trade-offs in sorghum. Genetics 218:iyab087. https://doi.org/10.1093/genetics/iyab087

NBPGR (2021) National Bureau of Plant Genetic Resources. http://exploreit.icrisat.org/profile/genetic%20resources/67

Ofosu FK, Elahi F, Daliri EB, Tyagi A, Chen XQ, Chelliah R, Kim JH, Han SI, Oh DH (2021) UHPLC-ESI-QTOF-MS/MS characterization, antioxidant and antidiabetic properties of sorghum grains. Food Chem 337:127788. https://doi.org/10.1016/j.foodchem.2020.127788

Olatoye MO, Marla SR, Hu Z, Bouchet S, Perumal R, Morris GP (2020) Dissecting adaptive traits with nested association mapping: genetic architecture of inflorescence morphology in sorghum. G3 10:1785–1796. https://doi.org/10.1534/g3.119.400658

Omrani M, Keshavarz M, Nejad Ebrahimi S, Mehrabi M, McGaw LJ, Ali Abdalla M, Mehrbod P (2020) Potential natural products against respiratory viruses: a perspective to develop anti-COVID-19 medicines. Front Pharmacol 11:586993. https://doi.org/10.3389/fphar.2020.586993

Ongom PO, Ejeta G (2018) Mating design and genetic structure of a multi-parent advanced generation intercross (MAGIC) population of sorghum (*Sorghum bicolor* (L.) Moench). G3 8: 331–341. https://doi.org/10.1534/g3.117.300248

Oria MP, Hamaker BR, Axtell JD, Huang CP (2000) A highly digestible sorghum mutant cultivar exhibits a unique folded structure of endosperm protein bodies. Proc Natl Acad Sci U S A 97: 5065–5070. https://doi.org/10.1073/pnas.080076297

Paiva CL, Evangelista WP, Queiroz VA, Gloria MB (2015) Bioactive amines in sorghum: method optimisation and influence of line, tannin and hydric stress. Food Chem 173:224–230. https://doi.org/10.1016/j.foodchem.2014.10.039

Przybylska-Balcerek A, Frankowski J, Stuper-Szablewska K (2018) Bioactive compounds in sorghum. Eur Food Res Technol 245: 1075–1080. https://doi.org/10.1007/s00217-018-3207-0

Przybylska A, Frankowski J, Stuper K (2019) Bioactive compounds in sorghum. Eur Food Res Technol 245:1075–1080. https://doi.org/10.1007/s00217-018-3207-0

Queiroz VAV, da Silva A, Beserra C, de Carvalho CWP, Liboreiro C, Costa P, Pereira RR (2018) A low calorie and nutritive sorghum powdered drink mix: influence of tannin on the sensorial and functional properties. J Cereal Sci 79:43–49. https://doi.org/10.1016/j.jcs.2017.10.001

Seyoum Y, Negussie R, Baye K (2016) Nutrient retention and fate of iron-binding phenolic compounds during the Injera processing of tannin-free and high-tannin sorghum. J Sci Food Agric 96:1541–1547. https://doi.org/10.1002/jsfa.7246

Shen S, Huang R, Li C, Wu W, Chen H, Shi J, Chen S, Ye X (2018) Phenolic compositions and antioxidant activities differ significantly among sorghum grains with different applications. Molecules 23:1203–1215. https://doi.org/10.3390/molecules23051203

Song G, Jia M, Chen K, Kong X, Khattak B, Xie C, Li A, Mao L (2016) CRISPR/Cas9: a powerful tool for crop genome editing. Crop J 4:75–82. https://doi.org/10.1016/j.cj.2015.12.002

Szerszunowicz I, Kłobukowski J (2020) Characteristics of potential protein nutraceuticals of plant origin with antioxidant activity. Molecules 25:1621. https://doi.org/10.3390/molecules25071621

Upadhyaya HD, Vetriventhan M, Deshpande S (2016) Sorghum germplasm resources characterization and trait mapping. In: Rakshit S, Wang Y (eds) The sorghum genome. Springer, pp 77–94

USDA (2021) Agricultural Research Service, National Plant Germplasm System, Germplasm Resources Information Network GRIN-Taxonomy. National Germplasm Resources Laboratory. https://npgsweb.ars-grin.gov/gringlobal/taxon/taxonomysimple

Von Born P, Bernardo-Faura M, Rubio-Somoza I (2018) An artificial miRNA system reveals that relative contribution of translational inhibition to miRNA-mediated regulation depends on environmental and developmental factors in Arabidopsis thaliana. PLoS ONE 13: e0192984. https://doi.org/10.1371/journal.pone.0192984

Wang L, Lu Z, Regulski M, Jiao Y, Chen J, Ware D, Xin Z (2021) BSAseq: an interactive and integrated web-based workflow for identification of causal mutations in bulked F2 populations. Bioinformatics 37: 382–387. https://doi.org/10.1093/bioinformatics/btaa709

Winn JA, Mason RE, Robbins AL, Rooney WL, Hays DB (2009) QTL mapping of a high protein digestibility trait in *Sorghum bicolor*. Int J Plant Genomics 2009:471853. https://doi.org/10.1155/2009/471853

Wongwaiwech D, Weerawatanakorn M, Boonnoun P (2020) Subcritical dimethyl ether extraction as a simple method to extract nutraceuticals from byproducts from rice bran oil manufacture. Sci Rep 10:21007. https://doi.org/10.1038/s41598-020-78011-z

Wong JH, Lau T, Cai N, Singh J, Pedersen JF William, Vensel H, Hurkman WJ, Wilson JD, Lemaux PG, Buchanan BB (2009) Digestibility of protein and starch from sorghum (Sorghum bicolor) is linked to biochemical and structural features of grain endosperm. J Cer Sci 49:73-82. https://doi.org/10.1016/j.jcs.2008.07.013

Wu Y, Yuan L, Guo X, Holding DR, Messing J (2013) Mutation in the seed storage protein kafirin creates a high-value food trait in sorghum. Nat Commun 4:2217. https://doi.org/10.1038/ncomms3217

Wu G, Bornman JF, Bennett SJ, Clarke MW, Fang Z, Johnson SK (2017) Individual polyphenolic profiles and antioxidant activity in sorghum grains are influenced by very low and high solar UV radiation and genotype. J Cereal Sci 77:17–23. https://doi.org/10.1016/j.jcs.2017.07.014

Xin Z, Wang ML, Barkley NA, Burow G, Franks C, Pederson G, Burke J (2008) Applying genotyping (TILLING) and phenotypinganalyses to elucidate gene function in a chemically induced sorghum mutant population. BMC Plant Biol 8:103. https://doi.org/10.1186/1471-2229-8-103

Xin Z, Wang M, Cuevas HE, Chen J, Harrison M, Pugh NA, Morris G (2021) Sorghum genetic, genomic, and breeding resources. Planta 254:114. https://doi.org/10.1007/s00425-021-03742-w

Yu J, Yan F, Lu Q, Liu R (2018) Interaction between sorghum procyanidin tetramers and the catalytic region of glucosyltransferases-I from *Streptococcus mutans* UA159. Food Res Int 112: 152–159. https://doi.org/10.1016/j.foodres.2018.06.027

Zhu H, Li C, Gao C (2020) Applications of CRISPR–Cas in agriculture and plant biotechnology. Nat Rev Mol Cell Biol 21:661–677. https://doi.org/10.1038/s41580-020-00288-9

Breeding Efforts on Grain Micronutrient Enhancement in Pearl Millet

Mahalingam Govindaraj and Mahesh Pujar

Contents

Abstract

The increasing population growth essentially requires more diverse food in the future. Pearl millet is a major source of diets that provides energy and nutrition for millions of people living in India and sub-Saharan Africa (SSA). Malnutrition prevalence is estimated at 2 billion people globally affected. Malnutrition is highly prevalent in India and SSA, and progress in addressing them through

M. Govindaraj (✉)
HarvestPlus Program, Alliance of Bioversity International and the International Center for Tropical Agriculture (CIAT), Cali, Colombia
e-mail: m.govindaraj@cgiar.org

M. Pujar
Crop Improvement Theme, Research Program-Asia, at ICRISAT, Hyderabad, India

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_7

various recurrent costing interventions is slow. Biofortification is a nutrition-defined cost-competitive and sustainable approach for combating malnutrition among resource-poor households in low- and middle-income countries. Enhancing grain iron (Fe) and zinc (Zn) contents is prioritized for major staples including pearl millet as these two micronutrients are predominately deficient in human populations. HarvestPlus-supported biofortification pearl millet breeding globally accomplished significantly higher levels for Fe (80 mg kg^{-1}) and Zn (60 mg kg^{-1}) in germplasm, breeding populations, and hybrid parents using conventional breeding approaches. To date, 12 biofortified hybrids are released and are benefiting more than 120,000 households. Genetic gain for Fe and Zn is gradually increased (42 mg kg^{-1} to >75 mg kg^{-1} Fe; 30 mg kg^{-1} to 50 mg kg^{-1} Zn) and is higher than yield. Average levels of Fe (42 mg kg^{-1}) and Zn (31 mg kg^{-1}) in commercial hybrids and increased climate variability for pearl millet-growing areas will affect nutrition levels besides diminishing yield potential. Therefore, breeding for yield and nutrition should go hand in hand with mainstreaming nutrition. The use of biofortified cytoplasmic male sterile (CMS) lines and restorers, breeding pipelines in crossing, and application of improved breeding methods through precision phenotyping, genomic selection, and speed breeding options can expedite mainstreaming progress in pearl millet. Bioavailability studies confirm the improved human health significance of biofortified varieties and hybrids. Accelerated public-private partnership is essential in achieving higher nutrition in competitively yielding hybrids. Consumer preferences for nutritious millet grains have increased, thus prospecting nutrient-dense varieties for improved human nutrition in India and SSA.

Keywords

Biofortification · Micronutrients · Iron · Zinc · Genomics · Mainstreaming

1 Introduction

Over 2 billion people globally are affected by at least one or more micronutrient deficiencies, which means that one-third of the global population is at risk of essential micronutrient deficiency (WHO 2019). The increasing population growth essentially requires nutrition-rich food besides energy in the future for their health and higher productivity. Pearl millet is a major source of diet that provided energy and nutrition for millions of people in India and sub-Saharan Africa (SSA) (Serba et al. 2020). Pearl millet is generally considered a nutrition-rich crop, but facts are rapidly changing owing to pearl millet breeding entirely focused on yield improvement in the last 30 years – will have a nutrient loss like other cereals. Malnutrition is preventable since it is the consequence of daily dietary nutrition deficiency. About 40% of Indian children under 5 are stunted, and 50% of women of reproductive age are anemic (NFHS, 2015–2016), whereas 24.1% of the population is malnourished in SSA (https://www.actionagainsthunger.org/africa-hunger-relief-facts-charity-aid).

Biofortification refers to a process of breeding staple food crops with higher micronutrient content in edible parts and is proven as an essential strategy in addressing malnutrition that is sustainable and cost-effective (Bouis et al. 2011). HarvestPlus program of the CGIAR supported many CGIAR and NARS centers for biofortification breeding targeting iron, vitamin A, and zinc with the main crops being beans, sweet potato, cassava, maize, rice, banana, wheat, and pearl millet (Lowe et al., 2022). The HarvestPlus-supported targeted biofortification breeding projects at ICRSIAT explored germplasm screening and investigated the feasibility of biofortification in pearl millet and aimed to support mainstreaming breeding and its goals with novel traits (Rai et al. 2014). ICRISAT pearl millet breeding program investigated the genetic variability of various grain quality traits (iron, zinc, ß-carotene) during the inception of biofortification projects. Preliminary research results revealed inadequate variability to undertake genetic improvements for ß-carotene; thus, focus was directed to Fe and Zn improvement (Rai et al. 2012b). The justification for the pearl millet biofortification initiative was very clear on millet consumer health and appropriate for additional resource mobilization for more nutrient-specific germplasm development and supporting variety/hybrid parent research in India and Africa. The major reasons for pearl millet biofortification investment are as follows: (i) more than 60% of the population in arid and semiarid tropics is malnourished, (ii) inadequate micronutrient levels among commercially grown pearl millet hybrids, and (iii) adequate genetic variation for micronutrients in breeding population and germplasm accessions. Pearl millet is cultivated close to 30 million ha globally covering five continents, viz., Asia, Africa, North America, South America, and Australia. Despite the fact that India (8 million ha) and Africa (about 18 million ha) contribute the majority of crop area (Yadav and Rai 2013). With 8.61 million tonnes, India is the world's largest producer of pearl millet (Directorate of Millets Development 2020). The major breeding goal of the ICRISAT pearl millet program is to (i) provide trait-specific germplasm and improved breeding lines and parents to NARS and other stakeholders; (ii) cultivate inter-institutional collaboration integrated conventional, participatory, and genomics-assisted breeding methods to develop widely adapted varieties and hybrid parents; and (iii) provide need-based capacity building on advanced tools and techniques. Although ICRISAT pearl millet breeding was established in the 1980s and the national program in the Indian Council of Agricultural Research (ICAR) in 1965 in India, no serious efforts were made to improve grain nutrition in pearl millet, while just a few attempts were made for screening germplasm before its inception of CGIAR biofortification program in 2005.

Screening efforts found highly significant and larger genetic variation among breeding populations (30–70 mg kg^{-1}), hybrid parents (25–90 mg kg^{-1}), and germplasm (28–120 mg kg^{-1}) at ICRISAT and NARS. On the other hand, negative correlations reported between micronutrient content and grain yield hindered the commercial breeding prospects. These studies' materials are highly selected for yield traits and not bred for micronutrients; therefore, there is no surprise in the observed trend. Some studies indicated that nonsignificant correlations bring hope for concurrent genetic improvement. In addition, a significant positive association between

Fe/Zn and grain yield has been reported suggesting the feasibility of breeding competitive biofortified varieties and hybrids in the future. Therefore, a biofortified cultivar with improved mineral content is easily accepted by consumers as it does not require a change in dietary habits. Biofortification programs in general target to improve the micronutrient content of those cultivars which have preferred agronomic and consumption traits like grain yield. HarvestPlus has initiated the development and promotion of many biofortified cultivars with improved micronutrient content across different food crops including pearl millet (Yadav et al. 2017). Systematic deliberation with ICAR, focusing on nutritional improvement besides yield improvement, was also mandatory in pearl millet. This landmark decision was taken by the pearl millet researchers to include iron and zinc concentrations as one of the promotion criteria for promoting entries in the national coordinated trials during the 52nd Annual Group Meeting of ICAR-AICRP on Pearl Millet at PAU, Ludhiana, on April 28, 2017 (Satyavathi et al. 2021). To date, 12 biofortified pearl millet cultivars have been released in India and West Africa (Govindaraj et al. 2019) and reached more than 120,000 households (https://www.cgiar.org/innovations/high-iron-pearl-millet-for-better-health/).

2 Pearl Millet and Selected Nutrition Traits

Pearl millet has significant potential as food, feed, and fodder crop in subsistence farming in the semiarid tropics. It can produce grains with high nutritive value even under hot, dry conditions on infertile soils of low water-holding capacity, where other cereal crops fail (Khairwal and Yadav 2005). The energy value of pearl millet grain is relatively higher compared to maize, wheat, or sorghum (Hill and Hanna 1990). Globally, pearl millet is cultivated on about 27 m ha of this, and India annually cultivates 9.3 m ha and produces about 9.5 m t. Pearl millet genetic improvement in India and SSA is critical for contributing to food production in semiarid tropic regions. There are many significant milestones achieved over the five decades. For instance, grain yield significantly surpassed from 4.5 kg/ha/yr (pre-green revolution [GR] era) to 31.1 kg/ha/yr (post-GR era) that accounts for a 188% increase in productivity since the green revolution (the 1970s) in India (Yadav et al. 2019). Crop improvement efforts in enhancing further grain and fodder with biotic and abiotic tolerances are being the topmost priority in both public and private sector breeding programs. The expansion of pearl millet uses to forage and dry fodders in drylands captivated the improvement in forage and fodder quantity per unit area. This supports food for humans and fodder for livestock in subsistence farming practices. Therefore, it is very clear that smallholder farming is dominant in semiarid regions and continues to be overwhelmed by numerous issues including poverty and malnutrition (Ryan and Spencer 2001; Sharma et al. 1996: Dar 2011). In semiarid tropic regions, household income and food expenditure play a significant role in determining family nutritional status (Padmaja et al. 2019). Addressing malnutrition in semiarid tropics where poverty and malnutrition persist together is not a supreme task of crop improvement history. The inception of biofortification

initiatives demonstrated the nutrition breeding feasibility and sustainability of nutrition supply to rural remote households. Biofortification breeding in pearl millet is a recent development and was led by ICRISAT in partnership with NARS in India and SSA.

In the search for nutritional functional properties from pearl millet grains, many traits are randomly reported in the past. However, their lower levels, less relevance to predominant nutrient deficiency among populations and nonsignificant genetic variations within crop genetic resources, diverted the breeding focus of Fe and Zn and some extent to proteins (Govindaraj et al. 2022). Pearl millet has an average of 11–12% proteins with moderately balanced amino acids (Shobana et al. 2013). The contribution of pearl millet to the total nutrient intake (Fe and Zn) from all foods widely varied across rural India; however, in some parts of rural India (Rajasthan, Maharashtra, and Gujarat), contribution of pearl millet to micronutrient (Fe and Zn) intake is higher. For instance, in these regions, pearl millet contributes 19–63% of the total Fe intake and 16–56% of the total Zn intake (Parthasarathy Rao et al. 2006). Therefore, it would be highly rewarding in providing additional Fe to their regular diet. Moreover, as the micronutrient requirements in human and plant nutrition are similar, genetic enhancement for grain minerals could improve human nutrition as well as farm productivity (Ma 2007). Today, addressing malnutrition is one of the priorities of the government of India, and pearl millet is one of the target crops in its NutriFarm initiative. Furthermore, a recent clinical study reveals that the bioavailability of both Fe and Zn from biofortified pearl millet is more than adequate to meet the physiological requirements for these micronutrients for children under 2 years old and young women (Kodkany et al. 2013; Cercamondi et al. 2013). Therefore, those farmers who grow biofortified pearl millet will have easy access to nutritious foods with minimal investments for their family's well-being, and surplus productions can be marketed for greater impact in millet-consuming communities. Therefore, to meet the daily energy and nutrition requirements, dietary diversity and pearl millet biofortification are very much closer and could prove to be an essential strategy for combating micronutrient malnutrition in India and SSA in the future.

3 Pearl Millet Nutrition Profile at a Glance

Overall mineral nutrition of pearl millet is generally higher than that of other major cereals such as rice, wheat, and maize (Adeola and Orban 1995). Deep looking into the differences among the commercial and breeding lines of pearl millet collection will be much higher and may exceed two- to three fold variation for most nutrition traits. A national collection of pearl millet varieties, hybrids, and key germplasm under ICAR-All India Coordinated Research Project on Pearl Millet (AICRP-PM) screened for major nutrition recently (Goswami et al. 2022). This collection consisted of 87 genotypes (mixture of varieties, hybrids, and germplasm). The starch content ranged from 50 to 63 g/100 g. Amylose content varied from 19 to 28 g/100 g; glucose and sucrose content varied in very narrow variation (<1 g/100 g). Protein content was found significantly higher (8–18 g/100 g). The anti-nutritional factors

like phytic acid content were found in a wider range (0.54–1.43 g/100 g), while the phenol content was a negligible amount (0.04–0.21 GAE g/100 g). However, the lipid variability is quite higher in this collection of genotypes (5–9%). About 20–32% of palmitic acid (C16:0), 3–8% of stearic acid (C18:0), 32–47% of linoleic acid (C18:2), and 22–33% of oleic acid (C18:1) were observed in the same collections (Goswami et al. 2022). On the other hand, high levels of grain minerals such as calcium, phosphorus, magnesium, manganese, zinc, iron, and copper in pearl millet are higher than in maize (Adeola and Orban 1995; Govindaraj et al. 2009). Values for different possible nutrients from pearl millet were listed in Table 1. Govindaraj et al. (2021) reported the wider genetic variation for P, K, Ca, and Mg content with a mean of 369, 489, 12, and 130 mg 100 g^{-1}, respectively, from large-scale field trials (14 trials) consisting of germplasm, breeding lines, commercial hybrids, and

Table 1 Nutrition profile of pearl millet foods reported in different sources

Nutrition profile	Pearl millet
Carbohydrates (Kcal)	67.5
Protein (g)	11.6
Fat (g)	5
Energy (kcal)	361
Crude fiber (g)	1.2
Mineral matter (g)	2.3
Ca (mg)	42
P (mg)	296
Fe (mg)	8
Mg (mg/100 g)	137
Na (mg/100 g)	10.9
K (mg/100 g)	307
Cu (mg/100 g)	1.06
Mn (mg/100 g)	1.15
Mb (mg/100 g)	0.069
Zn (mg/100 g)	3.1
Cr (mg/100 g)	0.023
Si (mg/100 g)	147
Cl (mg/100 g)	39
Thiamin (mg)	0.38
Niacin (mg)	2.8
Riboflavin	0.21
Vitamin A (carotene mg/100 g)	132
Folic acid (mg/100 g)	45.5
Vitamin B5 (mg/100 g)	1.09
Vitamin E (mg/100 g)	19

Source: Nutritive Value (macronutrients) of Indian Foods, NIN, Hyderabad, 2007
Source: Nutritive value (micronutrients and vitamins) of Indian foods, National Institute of Nutrition (2007); MILLET in your Meals, http://www.sahajasamrudha.org

advanced parents. The range of variability for these macronutrients across trials was 275–495 mg 100 g^{-1} for P, 340–725 mg 100 g^{-1} for K, 4–40 mg 100 g^{-1} for Ca, and 94–189 mg 100 g^{-1} for Mg. Similarly, the mean micronutrient contents across the 14 trials were 53 mg kg^{-1} for Fe, 41 mg kg^{-1} for Zn, and 13 mg kg^{-1} for both Mn and Na (Govindaraj et al. 2021). The magnitude of variability for macronutrients is higher than those observed for micronutrient contents with the following order K > P > Mg > Ca > Fe > Zn > Na > Mn. The results also revealed that the variability of P, K, Ca, Mg, Fe, Zn, Mn, and Na in parents/inbred trials was particularly larger than those found in the hybrids suggesting scope for improvement in commercial hybrids in the future otherwise underutilized this nutrition potential in presently grown pearl millet varieties and hybrids (Govindaraj et al. 2022). Therefore, untapped commercial feasibility prospects for genetic enhancement of these grain minerals in pearl millet along with productivity traits would greatly enhance pearl millet as the cheapest source of nutrition supply in drylands.

4 Pearl Millet Breeding at ICRISAT

Pearl millet remains one of the important staple crops in India after rice and wheat. The breeding program at ICRISAT generally develops open-pollinated varieties (OPVs) and hybrid parents (inbred lines) to produce hybrids which are the predominant cultivar type in India that covers about 5–6 m ha spread across three pearl millet cultivation zones (A, A_1, and B) of India. All these three zones are differentiated by a range of factors including plant types and environmental variables such as soil type, rainfall, and farm inputs. For example, zone A has >400 mm annual rainfall, rainy season cultivation, and sandy loam soil; zone A_1 has <400 mm rainfall, rainy season cultivation, dual-purpose OPVs and hybrids, and light sandy soils; zone B conquers with 400–700 mm rainfall and heavy and deep soils (AICRP-PM, 2020). Trait-based germplasms and elite hybrid parents are shared through the standard material transfer agreement (SMTA) process to public and private sectors to produce commercially viable OPVs and hybrids using different cytoplasmic male sterility lines (A_1, A_4, and A_5). The breeding program focuses largely on the rainy season practices in both A and B zones that require dual-purpose (grain and stover) materials, with disease tolerance to downy mildew and blast. The lead breeder has been in place since 2012/2013. Gujarat and South Rajasthan locations are used in the summer season for screening heat tolerance. The breeding is modernized and largely segmented into product profiles that target specific adaptation zone in India and SSA (https://repo.mel.cgiar.org/handle/20.500.11766/10236). The major contribution of ICRISAT and consortia is measured as more than 60% of released and commercially grown hybrids are either based on ICRISAT-bred parental lines or parental lines developed by public and private sectors from improved germplasm supplied from ICRISAT. In 2020, ICRISAT made major efforts on mainstreaming micronutrient traits in the breeding program through the product profile metrics and stage-gate pipeline development processes. Henceforth, the biofortification breeding pipelines are being merged into mainstream pre-breeding and product development.

5 Genesis of Targeted Breeding for Nutrition Traits

The growing evidence from HarvestPlus research supports the presence of a wide range of genetic variability for grain Fe and Zn contents and significant positive association between Fe and Zn and predominance of additive gene action for Fe and Zn content in pearl millet (Rai et al. 2012a; Govindaraj et al. 2013, 2019, 2020) indicating good prospects for concurrent genetic enhancement of Fe/Zn. Though an increased understanding of association and inheritance for Fe and Zn enhances breeding efficiency, however, there is less reliable information on the influence of yield on micronutrient content and vice versa. This helps to realize the level of genetic variation in grain Fe/Zn and any possible dilution and concentration effects associated with differences in the yield potential of the genotype. Studies on pearl millet (Rai et al. 2009), wheat (Morgounov et al. 2007; Shi et al. 2008; Garvin et al. 2006; Zhao et al. 2009), sorghum (Reddy et al. 2005), and maize (Bänziger and Long 2000) reported moderate to highly significant negative relationship between micronutrients and grain yield. This relation could also be affected by specific environmental conditions and particular genetic material under investigation (Bänziger and Long 2000; Brkic et al. 2004; Simic et al. 2009). Further studies need to elucidate the underlying genetic and physiological mechanisms of these relations. Towards this end, as part of the biofortification research program at ICRISAT, the present study was undertaken with the specific objectives to fulfill the breeding efficiency gaps; examine the magnitude and nature of genetic variance; combine ability, heterosis, and G $\times$ E interaction effects for Fe and Zn with agronomic traits including grain yield; and examine the relationship between Fe, Zn, 1000-grain mass, and grain yield. This information is pertinent for the development of micronutrient-dense genotypes and for devising a sound breeding approach for the improvement of Fe and Zn content without compromising the grain yield.

6 Requirement for Nutrition Breeding

The goal of plant breeding in the past and at present focuses on achieving higher yield per unit area. This has been achieved in most crops by various contributing traits such as biotic and abiotic stress tolerances. India achieved a historical food grain production record in 2021–2022 which is about 215 m t. On the other hand, the unfortunate situation of our major food crops is poor sources of micronutrients required for human well-being. To address the problem of malnutrition and staple crops' nutritional deficiency together, greater crop improvement emphasis is required with the committed vision of nutritional trait breeding besides yield improvement. Like other major crops, the pearl millet breeding program also targeted major constraints of abiotic and biotic stress tolerances and associated traits towards the ultimate goal of achieving higher productivity but not nutrition inclusive. Consequently, grain quality trait improvement is not a core breeding trait. In the interest of academic and public health, greater progress has been made in protein, iron, and zinc content improvement in the recent past by the CGIAR program in

association with NARS. It is well-known fact that it is impossible to achieve this vision in a singular crop or organization; therefore, countrywide policies and platforms should be created and liked to global networks, nutrition, and market standards. To date, no significant progress has been made and standardized breeding approaches for most required essential nutritional traits except Fe, Zn, and provitamin A (Govindaraj et al. 2019). Therefore, to meet the daily energy and nutrition requirements, dietary diversity and biofortification are very much closer than we looked in the past, and the crop improvement program should not ignore its developmental goals and investments. HarvestPlus program extends its partnership and capacity to help national and international public and private breeding sectors to set up the nutritional standards and phenotyping platforms that enable the breeding and selection of elite nutrition-rich varieties through national and regional testing networks. The major breeding requirements for nutritional traits are (i) the status of genetic variation available within the crop, (ii) a rapid phenotyping method, (iii) genetics of traits and traits associated with yield parameters, and (iv) breeding targets and publicly available standards for given nutrition traits. These requirements were briefly discussed in the following sections.

6.1 Genetic Variability

A preliminary assessment of the magnitude of genetic variability for grain Fe and Zn content in various pearl millet materials through a random sample of 120 entries that included a diverse range of hybrid parents improved populations, population progenies, and germplasm accessions (Fig. 1). Variability among these entries was 30–76 mg kg^{-1} Fe and 25–65 mg kg^{-1} Zn content (Velu et al. 2007). Recent studies

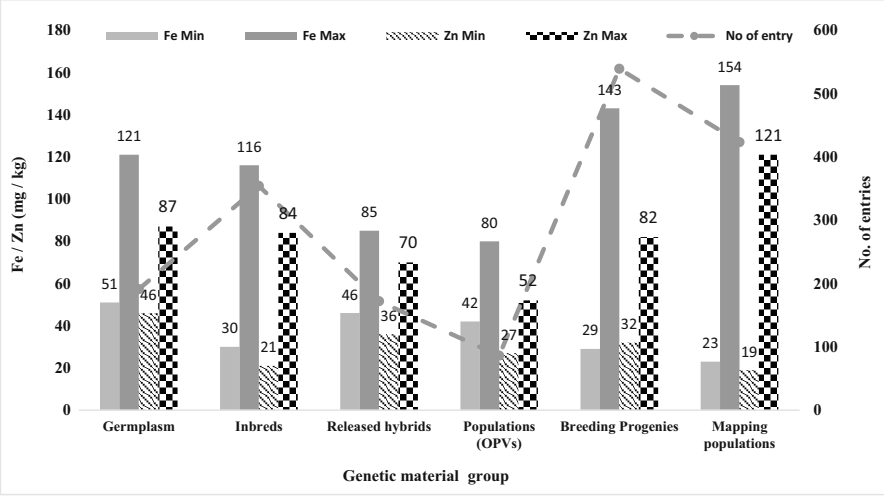


Fig. 1 Genetic variability for grain iron and zinc content in pearl millet

in the core collection germplasm accessions stored at ICRISAT differed significantly for these two micronutrients, especially with over two-fold variation for Fe (34–90 mg kg^{-1}), Zn (30–74 mg kg^{-1}), and Ca (85–249 mg kg^{-1}) (Govindaraj et al. 2020). Several field trials and research supported by the HarvestPlus Challenge Program of the CGIAR and conducted in partnerships with public and private sector research organizations have shown much large variability for Fe (30–120 mg kg^{-1}) and Zn (20–90 mg kg^{-1}) in the pearl millet germplasm and breeding lines suggesting good prospects of developing cultivars with higher levels of these micronutrients in the future without searching in unadopted germplasm as a trait source (Govindaraj et al. 2020; Rai et al. 2014).The screening of 297 Iniadi germplasm accessions from Western Africa (Togo, Eastern Ghana, Southern Burkina Faso, and Western Benin) has shown a wide variability for grain Fe and Zn content (Rai et al. 2008). Several studies in biofortification research at ICRISAT showed a wide range of variability for grain Fe and Zn densities in diverse breeding materials, for instance, Iniadi germplasm accessions (51–121 mg kg^{-1} Fe; 46–87 mg kg^{-1} Zn), population progenies (18.0–135.0 mg kg^{-1} Fe; 22.0–92.0 Zn), inbred parents (30.3–102.0 mg kg^{-1} Fe; 27.4 to 84.0 mg kg^{-1} Zn), hybrids derived from diverse inbreds (25.8–80.0 mg kg^{-1} Fe; 22.0–70 mg kg^{-1} Zn), and commercial hybrids (31.0–61.0 mg kg^{-1} Fe; 32.0–54.0 mg kg^{-1} Zn) (Govindaraj et al. 2019, 2021). Analysis of 281 advanced breeding lines exhibited substantial variability for Fe (35–116 mg kg^{-1}) and Zn (21–80 mg kg^{-1}) (Pujar et al. 2020). Overall, the mean of all the commercial hybrid grain samples assessed over sites in India indicated 42 mg kg^{-1} Fe and 31 mg kg^{-1} Zn (Govindaraj et al. 2019). Other nutrition ranges were described in Sect. 3.

6.2 Phenotyping for Grain Nutrition Traits

Sampling is a major step in nutrition estimation. A systematic study on pearl millet suggested that this crop is fortunate to qualify by the use of open-pollinated grain samples which would save time and resources (Govindaraj et al. 2011; Rai et al. 2015). More details of grain sampling are described by Govindaraj et al. (2020). In the past, two major methods were used for grain mineral estimation, namely, inductively coupled plasma optical emission spectrometry (ICP-OES) and atomic absorption spectrometer (AAS). These two were excellent tools for measuring the nutrition in plant samples; however, both are very destructive and sensitive and have higher recurring and sample cost that limits the affordability of the crop breeding program (Govindaraj and Rai 2016; Govindaraj et al. 2019). For easy breeding operation, resource efficiency, and preliminary screening purposes, HarvestPlus systematically worked with experts and manufacturers. With the great partnership and support from Waite Analytical Services (WAS) at the University of Adelaide, Flinders University, Australia, and Oxford instruments (presently the same model owned by Hitachi) brought up an innovative X-ray-based scanning tool so-called energy-dispersive X-ray fluorescence (ED-XRF). ICP and AAS methods cost 10–20 USD per sample (Rai et al. 2012b) and analyze not more than 6000 samples annually. Pearl millet breeding program handles 12,000–15,000 grain samples

every year. This would cost to program more than 120,000 USD per year in targeted biofortification breeding. Near-infrared reflectance spectrometer (NIRS), also a tool that can be used for many pearl millet grain samples, was analyzed. NIRS data are highly significant and has positive correlation with ICP values for Fe/Zn however, appeared less efficient (repeatability, low and high group distinguish) than ED-XRF in which the lines in the upper half of the Fe content were identified and lower groups were easily discarded. Selected samples at later breeding or testing stages will be screened using ICP or similar destructive methods for final validations and monitoring of any sample contaminations. After several investigations, the ED-XRF method was proven and recommended as a rapid and cost-effective screening tool for a large number of grain samples of pearl millet and other crops (rice, wheat, and sorghum). Results from this laboratory showed a highly significant and positive correlation ($r > 0.98$; Fig. 2) between ICP and XRF values, both for Fe and Zn content (Rai et al. 2012b; Paltridge et al. 2012; Govindaraj et al. 2016b). Setting up this small machine requires less space, with no recurring expenditure, and it provides nondestructive analysis of 250–300 samples per day at the cost of <2 USD/sample

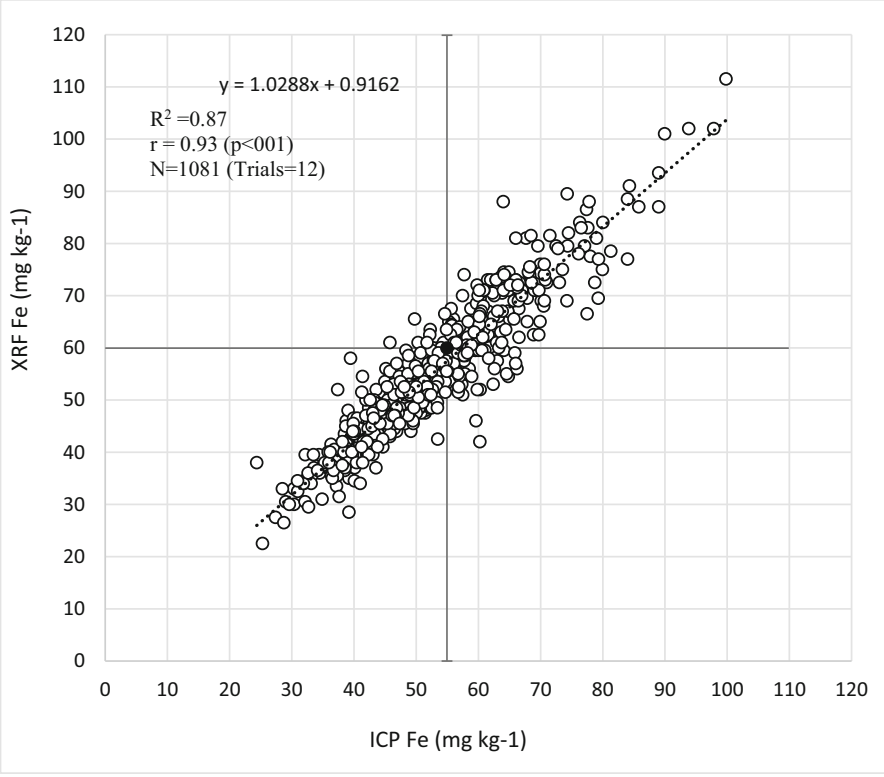


Fig. 2 Correlation between XRF and ICP data sets derived from 12 trials (1081 samples) of pearl millet

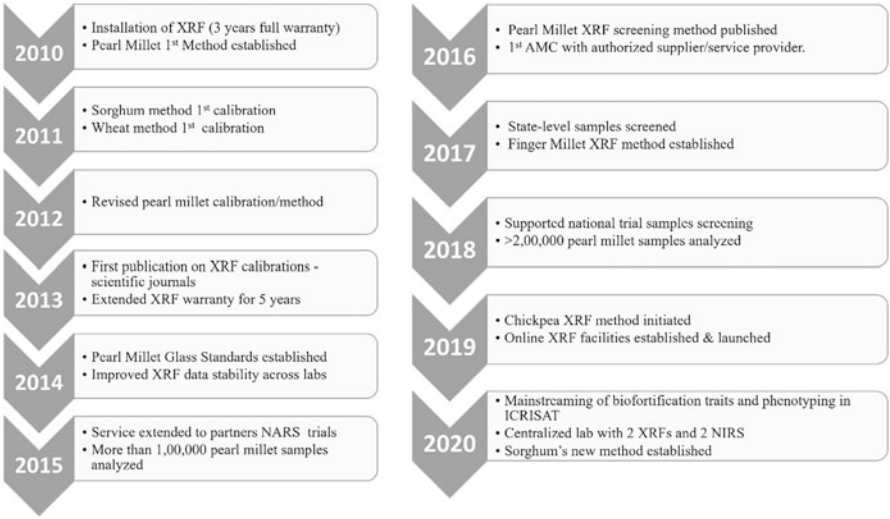


Fig. 3 Chronology of XRF tool and phenotyping facilities established at ICRISAT for pearl millet and other crops

for in-house materials. Based on these considerations, several XRF facilities were established across India (8–10 locations) and other countries (18–20 numbers) with support from HarvestPlus. Established XRF labs are key to success in accelerating and achieving the nutrition breeding targets in biofortification crops. Today, most crops have their specific standards operating screening procedures with glass standards which can ensure data quality and precision across laboratories. This tool is also being investigated for assessing other grain minerals (Ca, Se, P, and S) in the future. It is anticipated that the XRF tool can significantly reduce the cost and time not only for mineral analysis (Fe/Zn) but also for nutrition traits mainstreaming in regular breeding pipelines. Various stages of facility, development, and milestone of the XRF lab at ICRISAT are described in Fig. 3.

6.3 Nutrition Trait Genetics and Relationship with Agronomic Traits

Previous studies reported highly significant GCA and SCA effects for both Fe and Zn in pearl millet, indicating both additive and nonadditive genetic variances for controlling these two micronutrients. Both Fe and Zn contents in pearl millet are largely governed by additive genetic variance. However, all studies emphasized the magnitude of GCA variance, and its effect was much higher than SCA variance and its effect for both Fe and Zn, inferring the predominance of the additive gene for these trait expressions (Velu et al. 2011; Govindaraj et al. 2013; Kanatti 2014). In addition, other evidence indicating the grain Fe and Zn content in pearl millet is

under predominant additive genetic control included predictability ratios close to unity, a higher proportion of GCA-to-SCA ratio, and a considerably high and positive connection between the midparent value and the hybrid performance (Govindaraj et al. 2013). These findings suggest that there would be a minimal chance, if any, to take advantage of heterosis for these micronutrients in pearl millet, and breeding high Fe and Zn hybrids would require incorporating these traits into both parental lines. Since the additive genes with relatively lower G × E influence the accumulation of Fe and Zn in pearl millet grains (Govindaraj et al. 2016a; Kanatti 2014b), suggests high effectiveness of progeny selection in the pedigree selection or population breeding to develop lines and populations with increased levels of grain Fe and Zn contents. The higher additive genetic variance also prompts recurrent selection methods to improve the levels of grain Fe and Zn contents.

Besides gene action, a recent study on gene interactions performed using classical generation mean analysis revealed the existence of duplicate epistasis for grain Fe and Zn content in pearl millet and suggested avoiding the recombinant selection for high grain Fe and Zn content in early segregating generations (Boubacar Gaoh et al. 2020). The selection response and breeding behavior of the hybrid-parent breeding program are greatly influenced by knowledge of reciprocal cross differences and maternal effects. According to a study on the maternal effect on the accumulation of grain Fe and Zn content (Kanatti et al. 2018), the high-Fe trait can be incorporated into the genetic background of an elite line using high-Fe lines as either a female or male parent (considering within the seed and restorer gene pool) in a crossing program and selecting for elite agronomic performance with the high-Fe trait in the segregating populations. Furthermore, the genotype's innate ability to transmit traits from one generation to the next, or from parents to offspring, is known as heritability. Heritability is evaluated during selection and utilized as an indicator of trait transmission. Studies have shown the presence of high heritability for grain Fe and Zn contents in pearl millet indicating the genetic gain of these micronutrients can be increased by practicing simple selection.

The target attributes to be enhanced in pearl millet are grain Fe content, and Zn is improved as a related trait. This is made possible by the considerably extremely high positive correlation between Fe and Zn ($r = 0.43$ to 0.90, $P < 0.01$) (Rai et al. 2012b; Govindaraj 2012; Kanatti et al. 2016; Pujar et al. 2020). One of the important aspects while breeding for high grain Fe and Zn content is a non-compromise in the grain yield and its related attributes. Attributes such as grain size, flowering, and maturity are of economic significance that is of farmer's preference. Previous studies have shown the presence of moderate to high significant positive correlation for grain Fe and Zn content with 1000-grain weight in pearl millet (Velu et al. 2007, 2008a, b; Kanatti et al. 2014a; Pujar et al. 2020), which suggests that the improvement of Fe and Zn contents in grains would not affect the grain size. However, studies linking Fe and/or Zn with grain yield in pearl millet have found both positive and low to moderately negative associations (Gupta et al. 2009; Kanatti et al. 2014b; Rai et al. 2016; Pujar et al. 2020) and improvements for Fe and Zn in pearl millet will not reduce the grain yield since whenever the negative correlation has been detected, it is either very weak or found to be nonsignificant.

Below are key strategic findings and their breeding implications for nutrition traits

Strategic research findings	Breeding implications
• Fe and Zn are predominantly under additive genetic control and no-better-parent heterosis	• Need to breed for high Fe and Zn in both parental lines
• Mostly highly significant and positive association between line per se and general combining ability (GCA)	• Per se would be highly effective in selecting for GCA
• Large environment effect and large G × E interaction: G × E for Fe lower than for grain yield	• Need for multilocation testing through partnerships
• Cost-effective tool (XRF) for rapid analysis of Fe and Zn	• Facilitates handling of a large number of breeding lines (250–300 samples/day @ 2 USD)
• Large seed set effect, no xenia effect and dust contamination	• Open-pollinated seeds can be used for reliable estimation of Fe and Zn
• Fe and Zn highly significantly correlated, while Fe and Zn vs. grain yield mostly negatively correlated, but low to moderate and always not significant	• Simultaneous improvement for both micronutrients. High grain yield can be combined with high Fe/Zn by using large populations

6.4 Pearl Millet Breeding Priority and Product Profile

The crop breeding program has a historical target of improving grain yield per unit area through building genetically associated traits which directly or indirectly contribute to higher productivity. Breeding for nutritional traits is not an exception and has measurable as well as achievable target levels for given nutrition. Target levels were set to achieve an impact on health for the primary target population, particularly women of reproductive age and children (Pfeiffer and McClafferty 2007). Pearl millet is set to have a target increment of +30 mg kg^{-1} for Fe over the country- or region-specific baselines to achieve a required contribution to the estimated average requirement for Fe from the biofortified pearl millet. While a global baseline is estimated at 47 mg kg^{-1} and is systematically derived based on available commercial varieties at a given time, thus, in most cases, country-specific baselines may vary widely (Govindaraj et al. 2019; Govindaraj and Rai 2016). However, global breeding standards will serve as a unified reference for breeders, nutritionists, socio-economist, and market experts. Pearl millet breeding target increments (77 mg kg^{-1} Fe) are adjusted for its per capita consumption, bioavailability, and in many other processing-based crops, retention losses during processing, storage and cooking were also considered (Van Der Straeten et al. 2020). Pearl millet is generally consumed as whole grain (Govindaraj et al. 2020). WC C75 is the first ICRISAT-bred pearl millet variety which was widely grown for a long time in the 1980s and 1990s. In the initial years of biofortification research at ICRISAT, a 2-year study showed WC C75 having 42 mg kg^{-1} of Fe content (Velu et al. 2007). HarvestPlus set an ambitious working baseline of 47 mg kg^{-1} (5 mg kg^{-1}

above WC C75) and worked out a target level of 77 mg kg^{-1}, assuming 300 g/d of consumption, 90% retention, and 5% bioavailability. OPVs are the cultivars of the past, grown on <500,000 ha, while hybrids are cultivated on 4.5 million ha in India. Thus, the baseline of Fe was re-examined using more extensive hybrid trial data of this study. The relative performance of cultivars across the environment changes due to G × E interaction. However, relative means of various groups of cultivars, if the numbers in these groups are sufficiently large, are unlikely to be affected by G × E interaction. Thus, the relationship of the mean of 15 common hybrids evaluated in a two-location trial in 2011 and the ten-location trial in 2012 was used to assess the probable mean of the remaining 107 hybrids in 2012 based on their mean in 2011. And these values were used to estimate the baseline of Fe content. Product profile refers to the analysis of the resultant product to be generated through the breeding investment in terms of its performance, strength, and weakness in comparison to the already existing variety. Product profiles are defined based on the suggestions from market analysts, consumers, and multidisciplinary research and development teams. Product profile has two groups of traits: (i) must-have traits (present product preferences) and (ii) nice-to-have traits (future use). Fe and Zn are now transitioning from nice-to-have to must-have trait class in pearl millet for mainstreaming.

ICRISAT's pearl millet breeding in India has set up product profiles among which product profile 1 focuses on a 4.5 m ha area composed of A and B zones that include East Rajasthan, Central and South Gujarat, Haryana, Uttar Pradesh, Maharashtra, and Peninsular India (400–700 mm/annum). Under this, the area of prioritizations (Fig. 4) includes the development of parent lines of medium to late maturity

Trait priority		
Yield	Grain	*****
	Stover	***
Biotic Stress resistance	Downy mildew	*****
	Blast	**** (new)
	Ergot, Smut, Rust	**
Abiotic stress tolerance	Drought	** (strategic)
	Heat	** (new)
	Salinity	*
Nutrition	Iron and zinc	****(new)

** low priority and ***** highest priority*

Fig. 4 Prioritization of pearl millet genetic improvement and local testing zone in India

(70–90 days), dual-purpose (grain and fodder) nutritious (high Fe $\geq$60 mg kg^{-1} and Zn $\geq$ 35 mg kg^{-1}) hybrids with disease resistance (downy mildew and blast), for adaptation to better endowed environments. Product profile 2 focuses on Western Rajasthan and drier parts of Gujarat and Haryana. Under this, the area of prioritizations includes the development of parent lines with early maturity (65–75 days), dual-purpose nutritious (high Fe $\geq$50 mg kg^{-1} and Zn $\geq$35 mg kg^{-1}) hybrids with disease resistance for blast and downy mildew. In addition, it also focuses on the improvement in terminal drought tolerance. Figure 4 explains the pearl millet national testing, priority, and relevance for product release.

7 Status of Biofortified Hybrids and Performance

A trial consisting of 40 designated hybrid parents (20 each of seed parents and restorer parents), 30 each of improved populations and population progenies, and 20 germplasm accessions of diverse origin, was evaluated during the 2004 rainy and post-rainy seasons at ICRISAT, Patancheru. Based on the mean performance across the two seasons, 863B had the highest level of Fe (72.7 mg kg^{-1}) among the hybrid parents (Velu et al. 2007). A progeny from an open-pollinated variety released in India had 75.7 mg kg^{-1} Fe, which was the highest Fe level in the trial. These Fe levels are about twice those reported in wheat germplasm (Graham et al. 1999). It was also observed that 863B had 55.8 mg kg^{-1} Zn, ranking among the top five hybrid parents for this trait. The highest Zn level recorded for a hybrid parent in the trial was for seed parent 843B (59.6 mg kg^{-1}), while the highest Zn level for any entry in the trial was 63.7–64.8 mg kg^{-1} for the two progenies from AIMP 92901. Thus, 863B serves as an excellent source of high Fe and Zn contents in a commercial parental line. The salient features of some of the released biofortified pearl millet cultivars in India are presented in Table 2.

8 NARS Breeding Lines and Hybrids Characterized for Fe/Zn Content

The All India Coordinated Pearl Millet Improvement Project (AICPMIP) provided grain samples of hybrids tested in four different advanced hybrid trials in 2011. XRF analysis of grain samples of ten hybrids from advanced hybrid trial for A-zone (medium maturity) from Jodhpur, Rajasthan, showed the Fe content ranging from 22 to 41 mg kg^{-1} and Zn content ranging from 37 to 45 mg kg^{-1}. In the advanced hybrid trial of late maturity, the Fe content among the 16 hybrids varied from 31 to 43 mg kg^{-1} and Zn content from 33 to 43 mg kg^{-1}. In both trials, the highest Fe content was recorded for Pusa 23, a popular hybrid bred at IARI and released in 1987. The Fe content in 12 hybrids in the advanced hybrid trial for B-zone (medium maturity) at Dhule in Maharashtra varied from 43 to 62 mg kg^{-1}, and Zn content varied from 35 to 48 mg kg^{-1}. Pusa 23 and ICMH 356 (an ICRISAT hybrid released in 1993) had the highest and similar Fe content. Grain samples of 12 hybrids

Table 2 Performance and salient features of released biofortified pearl millet cultivars in India (Govindaraj et al. 2020)

Hybrid	Release year	Grain color	Grain size	Yield potential (t ha^{-1})[a]	Iron content (mg kg^{-1})[a]
Dhanashakti (variety)	2014	Dark gray	Bold	2.0	71
ICMH1202 (AHB1200Fe)	2017	Gray	Bold	3.5	70
ICMH 1203 (HHB 299)	2017	Gray	Bold	3.2	73
ICMH 1301 (DHBH1211)	2018	Gray	Bold	3.3	78
ICMH 1501 (HHB 311)	2018	Gray	Medium	3.5	70
ICMH 1502 (AHB1269)	2018	Gray	Bold	3.2	73
ICMH 1503 (RHB 233)	2018	Gray	Bold	3.2	80
ICMH 1504 (RHB 234)	2018	Gray	Medium	3.2	81

[a]Mean data from AICRP-PM test locations

evaluated in the advanced hybrid trial for B-zone (late maturity) were received from four locations (Dhule and Aurangabad in Maharashtra and Anantapur and Hyderabad in Andhra Pradesh). Based on the mean performance across four locations, the Fe content in these hybrids varied from 39 to 51 mg kg^{-1} and Zn content from 26 to 32 mg kg^{-1}. For the advanced hybrid trial for A-zone (late maturity) conducted in 2012, grain samples were received only from two locations. Based on the mean across the two locations, the Fe content in nine hybrids varied from 37 to 48 mg kg^{-1} and Zn content from 34 to 43 mg kg^{-1}. Grain samples of 40 breeding lines obtained from CCS Haryana Agricultural University (CCSHAU), Hisar, Haryana, and 157 breeding lines from Indian Agricultural Research Institute (IARI), New Delhi, were evaluated for Fe/Zn content. The XRF Fe content in CCSHAU lines varied from 23 to 112 mg kg^{-1}, and in IARI lines it varied from 39 to 117 mg kg^{-1}.

8.1 Commercial/Released Hybrids/OPVs Characterized for Fe/Zn Content

One hundred and twenty-eight hybrids (by name) developed by the public sector and private sector, which are released and/or under cultivation, were evaluated in 2011 under AICPMIP coordination at two diverse locations (Patancheru in peninsular India and Mandor in northern India). Based on the mean of the two locations, the ICP Fe content in these hybrids varied from 31 to 61 mg kg^{-1} and Zn content from 32 to 54 mg kg^{-1}. The highest-Fe entry in this trial was ICTP 8203, an OPV used as a high-Fe control. This variety had 73 mg kg^{-1} Fe and also the highest level of Zn

content (55 mg kg^{-1}). AICPMIP also coordinated the evaluation of 18 released OPVs at three locations in peninsular India and five locations in northern India. Based on the mean of eight locations, the ICP Fe content in these OPVs varied from 43 to 70 mg kg^{-1} and Zn content from 35 to 50 mg kg^{-1}, with ICTP 8203 having the highest levels of both micronutrients. Fifteen of the high-Fe hybrids ($\geq$50 mg kg^{-1}) for which the seeds could be available and five of the high-Fe OPVs ($\geq$53 mg kg^{-1}) identified in 2011 were re-evaluated under AICPMIP coordination at eight locations in 2012. Based on the mean performance across these locations, the XRF Fe content in the hybrids varied from 37 to 49 mg kg^{-1}. Three hybrids (86 M 86, Ajeet 38, and PAC 903) maintained their high rankings in both years. The Fe content in the OPVs varied from 39 to 64 mg kg^{-1}, with ICTP 8203 as the highest-Fe OPV. The XRF Zn content in hybrids varied from 34 to 46 mg kg^{-1}, while in the OPVs it varied from 38 to 44 mg kg^{-1}.

8.2 Elite Breeding Lines

HarvestPlus under the CGIAR consortium started the biofortification program at ICRISAT with the aim of developing high Fe and Zn biofortified hybrids and/or OPVs in pearl millet. Breeding for high Fe and Zn under the biofortification program could be able to develop several elite lines among which most of the lines were derived from the *iniadi* germplasm source. A study was conducted in which 281 advanced breeding lines (inbred lines) were evaluated for Fe and Zn contents in grains across two contrasting seasons (Pujar et al. 2020). These advanced breeding lines were collected from a diverse pool from the ICRISAT breeding program that constituted 112 restorer parents (R-lines), 110 seed parents (B-lines), 32 population progenies, and 27 germplasm progenies. The inbred performance for Fe varied from 35 mg kg^{-1} to 120 mg kg^{-1}, with an average grain Fe content of 75 mg kg^{-1}, whereas Zn varied from 19 mg kg^{-1} to 87 mg kg^{-1} with an average grain Zn content of 46 mg kg^{-1}. This accounts for three- to four-fold variability for grain Fe and Zn content among these advanced breeding lines/elite lines. Among 281 inbred lines, the top 15 elite lines with high grain Fe and Zn content identified are presented in Table 3.

9 G×E Effect on Grain Micronutrients

The grain Fe and Zn had no significant association with the available levels of these micronutrients in the soil (Govindaraj et al. 2019). In this context, it is important to note that the available Fe content at all the test locations and Zn content in most of the test locations were far above the critical levels (>2 mg kg^{-1}Fe and > 0.8 mg kg^{-1}Zn). The most extensive data on variability for grain Fe and Zn contents across the environments are available for Dhanshakti (Fig. 5), which was included as a control in several multilocation trials conducted over 3 years, giving 261 data points. Results showed that Fe content in the control variety can vary from

Table 3 Top 15 elite lines with high grain Fe and Zn contents identified in pearl millet

S. No	Elite line (B- and R-lines)	Fe (mg kg^{-1})	Zn (mg kg^{-1})
1	ICMB 100648	120	71
2	ICMR 1502	117	64
3	ICMP 100429	116	71
4	ICMR 100978	115	87
5	ICMR 100236	114	73
6	ICMR 100102	114	56
7	ICMP 100410	113	54
8	ICMB 100680	113	65
9	ICMP 100421	112	51
10	ICMP 100436	112	63
11	ICMB 100617	110	54
12	ICMR 100152	110	59
13	ICMB 100454	109	53
14	ICMP 100433	109	59
15	ICMR 100725	106	59

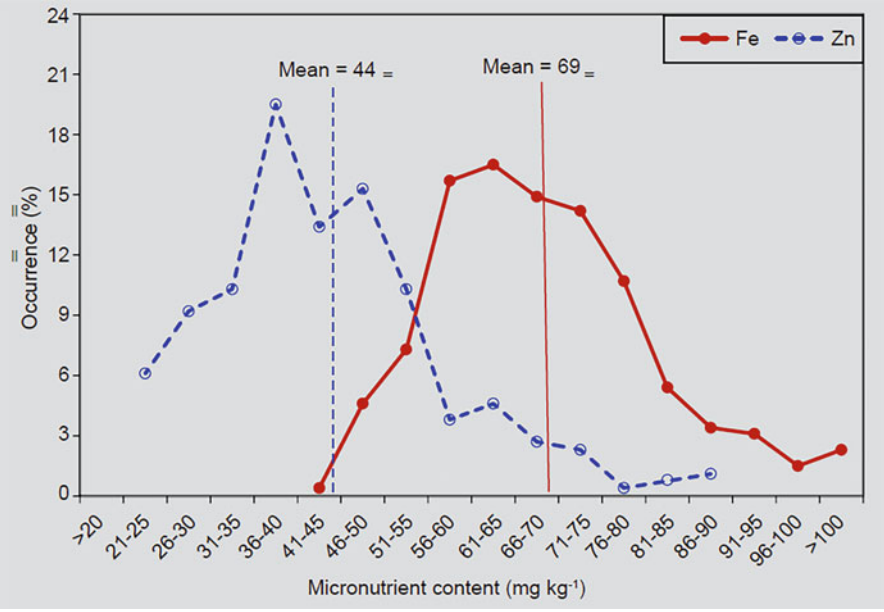


Fig. 5 Frequency distribution of Fe and Zn content in Dhanshakti across various trials pooled over location and year. (Source M Govindaraj)

40 to over 100 mg kg^{-1} and Zn content from 20 to 90 mg kg^{-1}. The mean Fe content was 69 mg kg^{-1}, ranging from 51 to 80 mg kg^{-1} in 80% of the cases, and the mean Zn content was 44 mg kg^{-1}, ranging from 26 to 55 mg kg^{-1}in 80% of the cases. This showed the extent to which Fe and Zn content can vary and the most likely levels of

these micronutrients expected in Dhanshakti. Similar variability patterns across the environments may be found in other genotypes. Significant G ×E interaction (genotype ×season) was noticed in the study conducted by Pujar et al. (2020) while evaluating 281 inbred lines for Fe and Zn contents in pearl millet grains. But it was evident that the magnitude of the effect of the environment was very less (sum of square analysis) than that due to differences caused by genotype alone. However, a trend was observed in the study wherein the magnitudes of the grain Fe content among inbreds were slightly higher in the summer season than in the rainy season, which might be due to the higher soil micronutrient absorption from the roots because of the high transpiration rate during the summer. However, the rankings of the inbreds across the seasons were consistent. A similar trend was also reported in earlier studies on pearl millet (Kanatti et al. 2014b; Govindaraj et al. 2013).

10 Conclusion

HarvestPlus-supported millet biofortification program led by ICRISAT established rapid screening platform, hybrid breeding target (72 mg kg^{-1} Fe; 45 mg kg^{-1} Zn), breeding strategy, and advanced high-Fe/Zn breeding lines that are being shared to NARS and private breeding programs. The dissemination of these nutrition-rich breeding lines and hybrid parents, and their utilization by user-research organizations (both public and private sectors) continuingly, as done so far for the non-biofortified materials, will make biofortified hybrid development a matter of routine and hence significantly contribute to improved nutrition. Mainstreaming of iron and zinc needs to be implemented in public and private sector breeding program to achieve the nutrition goal through strong national policies. Exploring other nutrition that is required for human and animal nutrition should be studied systematically, while the progress made in iron and zinc improvement will guide the process in the future. Additionally, the effects of the COVID-19 pandemic, coupled with the effects of ever-growing climate crises and food system conflicts, could exacerbate malnutrition, including micronutrient deficiencies, particularly in low- and middle-income countries (LMICs) and among young children and other vulnerable populations. Therefore, targeting most essential nutrients and vitamins including Fe and Zn besides yield and stress tolerance would be highly rewarding for millet scaling up and feeding millions of populations in semiarid tropics.

References

Adeola O, Orban JI (1995) Chemical composition and nutrient digestibility of pearl millet (*Pennisetum glaucum*) fed to growing pigs. J Cereal Sci 22:177–184

Bänziger M, Long J (2000) The potential for increasing the iron and zinc density of maize through plant breeding. Food Nutr Bull 21:397–400

Boubacar Gaoh BS, Gangashetty PI, Mohammed R, Dzidzienyo DK, Tongoona P (2020) Generation mean analysis of pearl millet [*Pennisetum glaucum* (L.) R. Br.] grain iron and zinc contents

and agronomic traits in West Africa. J Cereal Sci 96:103066. https://doi.org/10.1016/j.jcs.2020.103066

Bouis HE, Hotz C, McClafferty B, Meenakshi JV, Pfeiffer WH (2011) Biofortification: a new tool to reduce micronutrient malnutrition. Food Nutr Bull 32(suppl. 1):S31–S40

Brkic I, Simic D, Zdunic Z, Jambrovic A, Ledencan T, Kovacevic V, Kadar I (2004) Combining abilities of corn-belt inbred lines of maize for mineral content in grain. Maydica 48:293–297

Cercamondi CI, Egli IM, Mitchikpe E, Tossou F, Zeder C, Hounhouigan JD, Hurrell RF (2013) Total iron absorption by young women from iron-biofortified pearl millet composite meals is double that from regular millet meals but less than that from post-harvest iron-fortified. J Nutr 143:1376–1382

Dar WD (2011) Dryland agriculture in semi-arid tropics: constraints and opportunities. Indian J Dryland Agric Res Dev 26(1):1–7

Directorate of Millets Development (2020) Available online at: http://millets.dacfw.nic.in

Garvin DF, Welch RM, Finley JW (2006) Historical shifts in the seed mineral micronutrient concentration of US hard red winter wheat germplasm. J Sci Food Agric 86:2213–2220

Goswami S, Vinutha T, Tomar M, Sachdev A, Kumar RR, Veda K, Bansal N, Sangeetha V, Rama Prashat G, Singh SP, Khandelwal V, Kumar M, Satyavathi CT, Praveen S, Tyagi A (2022) Nutritional profiling of pearl millet genotypes. ICAR- AICRP-PM, Jodhpur

Govindaraj M (2012) Genetic detection of chromosomal interchanges and its consequence on plant breeding: a review. J Rec Adv Agri 1(3):63–68

Govindaraj M, Rai KN (2016) Breeding biofortified pearl millet cultivars with high iron density. Indian Farm 65(12):53–55. ISSN 0019-4786

Govindaraj M, Selvi B, Rajarathinam S (2009) Correlation studies for grain yield components and nutritional quality traits in pearl millet (*Pennisetum glaucum* (L.) R. Br.) Germplasm. World J Agric Res 5(6):686–689. ISSN 1817-3047

Govindaraj M, Kannan P, Arunachalam P (2011) Implication of micronutrients in agriculture and health with special reference to iron and zinc. Int J Agric Manag Dev 01(04):207–220. ISSN 2159-5852

Govindaraj M, Rai KN, Shanmugasundaram P, Dwivedi SL, Sahrawat KL, Muthaiah AR, Rao AS (2013) Combining ability and heterosis for grain iron and zinc densities in pearl millet. Crop Sci 53:507–517. https://doi.org/10.2135/cropsci2012.08.0477

Govindaraj M, Rai KN, Shanmugasundaram P (2016a) Intra-population genetic variance for grain iron and zinc contents and agronomic traits in pearl millet. Crop J 04(01):48–54. ISSN 2095-5421

Govindaraj M, Rai KN, Pfeiffer WH, Kanatti A, Shivade H (2016b) Energy-dispersive x-ray fluorescence spectrometry for cost-effective and rapid screening of pearl millet germplasm and breeding lines for grain iron and zinc density. Commun Soil Sci Plant Anal 47(18):2126–2134

Govindaraj M, Kanatti A, Rai KN, Satyanarayana T (2019) Does soil micronutrient variability in test locations influence performance of biofortified pearl millet in India? Better Crops South Asia 11:31–35. Available online at: http://oar.icrisat.org/11285

Govindaraj M, Virk PS, Kannati A, Cherian B, Rai KN, Anderson MS et al (2020) Biofortified pearl millet cultivars offer potential solution to tackle malnutrition in India. In: Kang MS (ed) Quantitative genetics, genomics and plant breeding. CAB International, Oxon, pp 385–396. https://doi.org/10.1079/9781789240214.0385

Govindaraj M, Kanatti A, Rai KN, Pfeiffer WH, Shivade H (2021) Association of grain iron and zinc content with other nutrients in pearl millet germplasm, breeding lines, and hybrids. Front Nutr 8:746625. https://doi.org/10.3389/fnut.2021.746625

Govindaraj M, Kanatti A, Rai KN, Pfeiffer WH, Shivade H (2022) Association of grain iron and zinc content with other nutrients in pearl millet germplasm, breeding lines, and hybrids. Front. Nutr 8:1357. https://doi.org/10.3389/fnut.2021.746625

Graham RD, Senadhira D, Beebe S, Iglesias C, Monasterio I (1999) Breeding for micronutrient density in edible portions of staple food crops: conventional approaches. Field Crops Res 60:57–80

Gupta SK, Velu G, Rai KN, Sumalini K (2009) Association of grain iron and zinc content with grain yield and other traits in pearl millet [*Pennisetum glaucum* (L.) R.Br.]. Crop Improv 36:4–7

Hill GM, Hanna WW (1990) Nutritive characteristics of pearl millet grain in beef cattle diets. J Anim Sci 68:2061–2066

Kanatti A (2014) Genetic architecture of grain iron and zinc densities and their association with agronomic traits in pearl millet (*Pennisetum glaucum* (L.) R. Br.). PHD thesis, Professor Jayashankar Telangana State Agricultural University, Hyderabad. Item availability restricted

Kanatti A, Rai KN, Radhika K, Govindaraj M, Sahrawat KL, Rao AS (2014a) Grain iron and zinc density in pearl millet: combining ability, heterosis and association with grain yield and grain size. Springerplus 3(763):1–12

Kanatti A, Rai KN, Radhika K, Govindaraj M, Sarawat LL, Srinivasu K, Shivade H (2014b) Relationship of grain iron and zinc content with grain yield in pearl millet hybrids. Crop Improv 41(1):91–96

Kanatti A, Rai KN, Radhika K, Govindaraj M (2016) Genetic architecture of open-pollinated varieties of pearl millet for grain iron and zinc densities. Indian J Genet Plant Breed 76(3): 299–303

Kanatti A, Govindaraj M, Rai KN, Rao AS (2018) Maternal effect on grain iron and zinc densities in pearl millet. In: 1st National Genetics Congress on Genetics for Sustainable Food, Health and Nutrition Security, December 14–16, 2018. ICAR- Indian Agricultural Research Institute, New Delhi

Khairwal IS, Yadav OP (2005) Pearl millet (*Pennisetum glaucum*) improvement in India – retrospect and prospects. Indian J Agric Sci 75:183–191

Kodkany BS, Bellad RM, Mahantshetti NS, Westcott JE, Krebs NF, Kemp JF, Hambidge KM (2013) Biofortification of pearl millet with iron and zinc in a randomized controlled trial increases absorption of these minerals above physiological requirements in young children. J Nutr 143:1489–1493

Low J, Anna-Marie B, Ilona P, Ekesa B, Heck S, Pfeiffer W (2022) Scaling Readiness of Biofortified Root, Tuber, and Banana Crops for Africa. In: Thiele G et al (eds), Root, Tuber and Banana Food System Innovations. pp 513–554. https://doi.org/10.1007/978-3-030-92022-7_17

Ma G (2007) Iron and zinc deficiencies in China: existing problems and possible solutions, PhD. Thesis, Wageningen University, The Netherlands

Morgounov A, Gomez-Becerra HF, Abugalieva A, Dzhunusova M, Yessimbekova M, Muminjanov H, Zelenskiy Y, Ozturk L, Cakmak I (2007) Iron and zinc grain density in common wheat grown in central Asia. Euphytica 155:193–203

NFHS (2015–2016) National Family Health Survey (NFHS-4). National Family Health Survey, India

Padmaja R, Soumitra P, Pingali P, Bantilan C, Kasala K (2019) Understanding nutritional outcomes through gendered analysis of time-use patterns in semi-arid India. Glob Food Sec 23:49–63. https://doi.org/10.1016/j.gfs.2019.04.001

Paltridge NG, Palmer LJ, Milham PJ, Guild GE, Stangoulis JCR (2012) Energy-dispersive X-ray fluorescence analysis of zinc and iron concentration in rice and pearl millet grain. Plant Soil 361: 251–260

Parthasarathy Rao P, Birthal BS, Reddy Belum VS, Rai KN, Ramesh S (2006) Diagnostics of sorghum and pearl millet grains-based nutrition in India. Int Sorghum Millets Newsl 47:93–96

Pfeiffer WH, Mcclafferty B (2007) Biofortification: breeding micronutrient-dense crops

Pujar M, Govindaraj M, Gangaprasad S, Kanatti A, Shivade H (2020) Genetic variation and diversity for grain iron, zinc, protein and agronomic traits in advanced breeding lines of pearl millet [*Pennisetum glaucum* (L.) R. Br.] for biofortification breeding. Genet Resour Crop Evol (TSI):1–14

Rai KN, Gowda CLL, Reddy BVS, Sehgal S (2008) The potential of sorghum and pearl millet in alternative and health food uses. Compr Rev Food Sci Food Saf 7:340–352. https://doi.org/10.1111/j.1541-4337.2008.00049.x

Rai KN, Gupta SK, Bhattachajee R, Kulkarni VN, Singh AK, Rao AS (2009) Morphological characteristics of ICRISAT-bred pearl millet hybrid parents. In: International Crops Research for the semi-arid tropics (Patancheru), p 176

Rai KN, Yadav OP, Gupta SK, Mahala RS, Gupta SK (2012a) Emerging research priorities in pearl millet. J SAT Agric Res 10:1–4. http://ejournal.icrisat.org/index.htm

Rai KN, Govindaraj M, Rao AS (2012b) Genetic enhancement of grain iron and zinc content in pearl millet. Qual Assur Saf Crops Foods 4:119–125. https://doi.org/10.1111/j.1757-837X.2012.00135.x

Rai KN, Gupta SK, Sharma R, Govindaraj M, Rao AS, Shivade H et al (2014) Pearl millet breeding lines developed at ICRISAT: a reservoir of variability and useful source of non-target traits. SAT eJournal 1:1–13. Available online at: http://oar.icrisat.org/id/eprint/8629

Rai KN, Velu G, Govindaraj M, Upadhyaya HD, Rao AS, Shivade H, Reddy KN (2015) Iniadi pearl millet germplasm as a valuable genetic resource for high grain iron and zinc densities. Plant Genet Res 13(01):1–8. ISSN 1479-2621 Item availability restricted

Rai KN, Yadav OP, Govindaraj M, Pfeiffer WH, Yadav HP, Rajpurohit BS, Patil HT, Kanatti A, Rathore A, Rao AS, Shivade H (2016) Grain iron and zinc densities in released and commercial cultivars of pearl millet *(Pennisetum glaucum)*. Indian J Agr Sci 86(3):291–296

Reddy BVS, Ramesh S, Longvah T (2005) Prospects of breeding for micronutrients and β-carotene-dense sorghums. Int Sorghum Millets Newsl 46:10–14

Ryan JG, Spencer DC (2001) Future challenges and opportunities for agricultural R & D in the semi-arid tropics. International Crops Research Institute for the semi-arid tropics

Satyavathi, CT, Ambawat S, Khandelwal V, Govindaraj M, Neeraja CN (2021) Micronutrient rich pearl millet for nutritionally secure India. ICAR- ALL India Coordinated Research Project on Pearl Millet, Jodhpur 18 pp

Serba DD, Yadav RS, Varshney RK, Gupta SK, Mahalingam G, Srivastava RK et al (2020) Genomic designing of pearl millet: a resilient crop for arid and semi-arid environments. In: Kole C (eds) Genomic designing of climate-smart cereal crops. Springer, Cham, pp 221–286. https://doi.org/10.1007/978-3-319-93381-8_6

Sharma MM, Garcia A, Qureshi BLR (1996) Overcoming malnutrition: is there an ecoregional dimension? vol 10. International Food Policy Research Institute

Shi R, Li H, Tong Y, Jing R, Zhang F, Zou C (2008) Identification of quantitative trait locus of zinc and phosphorus density in wheat (*Triticum aestivum* L.) grain. Plant Soil 306:95–104

Shobana S, Krishnaswamy K, Sudha V, Malleshi NG, Anjana RM, Palaniappan L, Mohan V (2013) Finger millet (Ragi, *Eleusine coracana* L.). Review of its nutritional properties, processing and plausible health benefits. Chapter 1. Adv Food Nutr Res 69:1–39. https://doi.org/10.1016/B978-0-12-410540-9.00001-6. PMid:23522794

Simic D, Sudar R, Ledencan T, Jambrovic A, Zdunic Z, Brkic I, Kovacevic V (2009) Genetic variation of bioavailable iron and zinc in grain of a maize population. J Cereal Sci 50:392–397

Van Der Straeten D, Bhullar NK, De Steur H, Gruissem W, MacKenzie D, Pfeiffer W, Qaim M, Slamet-Loedin I, Strobbe S, Tohme J, Trijatmiko KR, Vanderschuren H, Van Montagu M, Zhang C, Bouis H (2020) Multiplying the efficiency and impact of biofortification through metabolic engineering. Nat Commun 11(1):5203. https://doi.org/10.1038/s41467-020-19020-4

Velu G, Rai KN, Murlidharan V, Kulkarni VN, Longvash T, Raveendran TS (2007) Prospects of breeding biofortified pearl millet with high grain iron and zinc content. Plant Breed 126: 182–185

Velu G, Rai KN, Sahrawat KL, Sumalini K (2008a) Variability for grain iron and zinc contents in pearl millet hybrids. J SAT Agric Res 6:1–4

Velu G, Rai KN, Sharawat KL (2008b) Variability for grain iron and zinc content in a diverse range of pearl millet populations. Crop Improv 35(2):186–191

Velu G, Rai KN, Muralidharan V, Longvah T, Crossa J (2011) Gene effects and heterosis for grain iron and zinc density in pearl millet (*Pennisetum glaucum* (L.) R. Br). Euphytica 180:251–259. https://doi.org/10.1007/s10681-011-0387-0

World Health Organization [WHO] (2019) The state of food security and nutrition in the world 2019. Safeguarding against economic slowdowns and downturns. FAO, Rome

Yadav OP, Rai KN (2013) Genetic improvement of pearl millet in India. Agric Res 2:275–292. https://doi.org/10.1007/s40003-013-0089-z

Yadav OP, Upadhyaya HD, Reddy KN, Jukanti AK, Sushil P, Tyagi PK (2017) Genetic resources of pearl millet: status and utilization. Indian J Plant Genet Res 30(1):1–7

Yadav OP, Singh DV, Dhillon BS, Mohapatra T (2019) India's evergreen revolution in cereals. Curr Sci 116:1805–1808. https://doi.org/10.18520/cs/v116/i11/1805-1808

Zhao FJ, Su YH, Dunham SJ, Rakszegi M, Bedo Z, McGrath SP, Shewry PR (2009) Variation in mineral micronutrient concentrations in grain of wheat lines of diverse origin. J Cereal Sci 49: 290–295

Nutraceutomics of Foxtail Millet (*Setaria italica* L.): Insights

Jyothish Madambikattil Sasi, Paramananda Barman, and Charu Lata

Contents

Abstract

Setaria italica (L.), commonly known as foxtail millet, is a C_4 model cereal with a genome sequence that is compact and completely annotated. In many parts of Asia and Africa, it serves as a primary food and feed source. In the present world, where urbanization and farmland scarcity is drastically increasing, there is an urgent need to switch to cereal crops that can be grown with less inputs, consume limited resources, withstand adverse weather conditions, and provide the majority of the human diet's essential nutrients. Under these conditions, foxtail millet is an ideal cereal crop that has the potential to significantly contribute to international efforts to improve food and nutrition security. Foxtail millet grains, which are gluten-free, are an excellent dietary staple because of their high protein, fiber, carbohydrate, calcium, zinc, iron, vitamin, and lipid content. The therapeutic

J. M. Sasi · P. Barman · C. Lata (✉)
Inclusive Health and Traditional Knowledge Studies Division, CSIR-National Institute of Science Communication and Policy Research, New Delhi, India
e-mail: charulata@niscpr.res.in

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_8

benefits of foxtail millet include antihyperglycemic, antioxidative, anti-hyperlipidemic, anti-inflammatory, and antihypertensive activities and hence are strongly recommended that this grass family member be incorporated into a person's diet on a regular basis. The purpose of this chapter is to provide an overview of the nutraceutomics of the foxtail millet and the molecular efforts employed to improve the nutritional quality along with the traditional culinary applications of foxtail millet.

Keywords

Foxtail millet · Nutraceutomics · Nutrients · Omics · *Setaria* · Traditional food

1 Introduction

Millets are Poaceae family grass crops that have been cultivated for centuries. The Poaceae family consists of grasses and domesticated cereals including rice, maize, wheat, barley, millets, etc. "Millet" refers to cereal plants that produce numerous, minute seeds. Millet grains are a primary food source for many people worldwide, especially in the dry and semidry parts of Africa and Asia. Many economically relevant millets are used to produce biofuels and bioenergy products. Millets are particularly significant due to their high nutritional value and agro-industrial significance (Yousaf et al. 2021). After wheat, maize, sorghum, rice, and barley, millets are the world's sixth most important cereal grain crop. According to FAOSTAT 2020, between 2003 and 2018, nearly equal amounts of millets were produced in Asia and Africa, accounting for 96% of worldwide millet production, with the remaining 4% coming from the rest of the world (Yousaf et al. 2021). *Setaria italica* (L.) P. Beauv, commonly known as foxtail millet, is the world's second most extensively cultivated millet crop after pearl millet. Over 8000 years ago, in northern China, foxtail millet was said to have been domesticated from a wild species of green foxtail. Foxtail millet is a member of the *Setaria* genus, which consists of over 125 species predominantly spread in warm and temperate parts of the world (Lata et al. 2013). A typical foxtail millet plant attains a height of 120–200 centimeters and produces up to 13,000 seeds per plant. A thin stem terminates in a panicle of reddish or purple-hued bristles of around 5–30 cm in length. The characteristic shape of the panicle resembles a fox's tail, hence the common name "foxtail" for most cultivated *Setaria* species. *Setaria* genus members (foxtail millet [*S. italica*] and green foxtail [*S. viridis*]) are closely related to several millets, cereals, and biofuel grasses. Foxtail millet possesses a variety of beneficial characteristics, such as high photosynthetic efficiency, effective use of water and nitrogen even under unfavorable conditions, a high yield and productivity even with limited input resources, etc. In addition, foxtail millet grains are an excellent dietary staple because of their gluten-free, high protein, fiber, carbohydrate, calcium, zinc, iron, vitamin, and lipid content. The 2012 release of the *S. italica* genome sequence has substantially aided the development of large-

scale genomic resources (Bennetzen et al. 2012; Wang et al. 2012). Foxtail millet is a promising model for functional genomic studies in millets, cereals, and bioenergy grasses due to its small diploid genome (515 Mb), self-fertilization, and short cycle lengths (Lata et al. 2013; Lata and Shivhare 2017). Furthermore, because it is capable of NADP malic enzyme-type C_4 photosynthesis, it has become a model crop plant for genetic studies.

According to traditional Chinese medicine, consuming foods with pharmacological properties, such as foxtail millet, can reduce the chance of developing chronic diseases or even heal them. Foxtail millet is particularly well known in China for its beneficial effects on the digestive tract of humans. Recent research studies on foxtail millet have highlighted the anti-inflammatory, hyperglycemic, antihypertensive, and antihyperlipidemic potential (Akoh and Min 2007; Chow 2008; Hou et al. 2018; Jali et al. 2012; Li et al. 2022). Selection and, to a lesser extent, recombination breeding have primarily been used to improve foxtail millet crops (Hariprasanna et al. 2017). This crop's unique floral morphology and flowering behavior make it challenging to pursue crosses between the ideal parents. Therefore, foxtail millet has a limited potential for conventional breeding. Molecular marker-based, population studies and multi-omics approaches are needed to improve the breeding of foxtail millet.

2 Nutritional Composition and Therapeutic Values of Foxtail Millet

2.1 Nutritional Composition

The foxtail millet has more nutrients than other cereals like wheat and rice (Bandyopadhyay et al. 2017). Millet grains are a good source of nutrients, have a low glycemic index, are low in phytic acid, and do not contribute to stomach acid production. Millet is often referred to as "nutricereal" because it contains more nutrients per gram than other cereals. These include phytochemicals, micronutrients, antioxidants, and essential amino acids (except lysine and methionine) (Jali et al. 2012; Singh and Raghuvanshi 2012). According to some sources, foxtail millet (cultivar 'RAU-8') possesses the highest level of seed proteins among all millets (Chandel et al. 2014). In addition, foxtail millet ranks among the millets with the highest contents of protein, fat, ash, crude fiber, thiamine, and riboflavin (Bandyopadhyay et al. 2017).

Amino acids, both common and essential, have their foundation in proteins. Seed storage proteins (SSPs) found in millet grains are vital in facilitating seed germination after embryo development is complete. However, sulfur-containing amino acids like methionine and cysteine are more abundant in millets. All primary essential amino acids are present in foxtail millet except tyrosine (Table 1). The predominant amino acid found in foxtail millet is leucine (1040 mg/g of protein) followed by isoleucine (480 mg/g of protein) and valine (430 mg/g of protein). Foxtail millet also has the highest concentration of leucine and isoleucine of any millets (Bandyopadhyay et al. 2017). The foxtail millet has more elevated amounts of

Table 1 Nutrient composition of foxtail millet (Saleh et al. 2013; Bandyopadhyay et al. 2017)

Nutritional composition (per 100 g)	
Protein (g)	12.3
Fat (g)	4.3
CHO(g)	60.9
Crude fiber (g)	8
Energy (KJ)	331
Vitamin profile (mg)	
Thiamin	0.59
Riboflavin	0.11
Niacin	3.2
Folic acid	15
Vit A	32
Vit E	31
Vit B6	–
Vit B5	0.82
Essential amino acid profile (mg/g of N)	
Arginine	220
Histidine	130
Lysine	140
Valine	430
Tryptophan	60
Phenylalanine	420
Methionine	180
Threonine	190
Leucine	1040
Cysteine	100
Isoleucine	480
Tyrosine	–
Minerals profile (mg/100 g)	
Calcium	31
Iron	2.8
Magnesium	81
Potassium	250
Sodium	4.6
Sulfur	171
Copper	1.4
Molybdenum	0.07
Manganese	0.6
Zinc	2.4
Chlorine	37
Chromium	0.03
Fatty acid composition	
Palmitoleic	–
Palmitic	6.4
Oleic	13
Linolenic	–
Linoleic	66.5
Stearic	6.3

almost all essential amino acids than rice and wheat. Vitamins, an externally supplied dietary requirement, play a crucial role in maintaining human physiological homeostasis. Foxtail millet grains, like other types of millets, are an excellent source of a wide variety of vitamins (Table 1). The crop has more significant amounts of all vitamins except vitamin B6. It has the most significant concentrations of thiamine and vitamin E among millets as well as rice and wheat (Bandyopadhyay et al. 2017). The vitamin profiling of foxtail millet revealed that it is a good source of vitamins A and E and folic acid. In addition, foxtail millet is a good source of calcium, iron, magnesium, potassium, etc. Analysis of the mineral composition of foxtail millet revealed an abundance of potassium, followed by sulfur, magnesium, chlorine, and calcium (Table 1). According to the reports, foxtail millet is one of the richer sources of zinc and iron, with levels of 2.4 and 2.8 mg/100 g of grain, respectively. Foxtail millet generally has consistently higher contents of significant minerals and can serve as an effective supplement to our conventional dietary practices. Moreover, foxtail millet is an excellent source of fatty acids such as stearic and linoleic acids, both found in relatively low concentrations in other millet varieties and in cereals that are not classified as millet.

2.2 Biochemical Pathways Highlighted in Foxtail Millet

It was shown that the foxtail millet cultivated in various locales had significantly variable metabolites, likely due to the weather conditions in those areas (Yang et al. 2021a, b). Foxtail millet's secondary metabolite profile revealed an excess of flavonoids, especially in their glycosylated forms (Li et al. 2018). This indicates the enrichment of the flavonoid biosynthetic pathway in foxtail millet. Flavonoids have a multitude of beneficial medical effects, including anti-inflammatory, antiviral, and anticancer activities. Moreover, they shield the brain and heart from damage (Ullah et al. 2020). In contrast to other cereals, foxtail millet has been shown to have a significantly higher concentration of flavone O-aglycones, which may be the product of an enriched gene duplication event. A similar accumulation of phenolamides, which are 100 times more abundant in foxtail millet than in rice, has been documented. Phenolamides have long been crucial in various biological functions, such as plant growth and development, resistance to pathogens, and protection from abiotic stressors (Li et al. 2018). Phenolamides have been found to have beneficial benefits against metabolic syndrome and neurological illnesses, in addition to their antioxidant, anti-inflammatory, anticancer, and antibacterial capabilities (Roumani et al. 2020).

2.3 Therapeutic Uses

Localized inflammation is a defensive reaction involving several cells and chemicals. It is often associated with excessive secretion of pro-inflammatory cytokines, such as IL-1b, IL-6, or TNF-a, which can cause harm to the immune

system and hamper organ function, in addition to inducing or aggravating several disorders (Favalli 2020; Guo et al. 2020; Navarro-González and Mora-Fernández 2008). Anti-inflammatory testing using grain extracts from several foxtail millet accessions showed that most foxtail millet grain extracts inhibit the generation of pro-inflammatory mediators of macrophage response (Li et al. 2022). This highlights the possibility of using foxtail millet as an anti-inflammatory agent. It is well documented that millets are an excellent source of carbohydrates, which are needed for the health and operation of the digestive tract (Flight and Clifton 2006). They give the body the fuel to function and aid in delivering vital micronutrients (Eastwood 2003). Millets, particularly foxtail millet, are low in saturated fat; nonetheless, they are an excellent source of polyunsaturated fats, which help in reducing low-density lipoprotein (LDL) cholesterol (Akoh and Min 2007). As a result, having low levels of LDL cholesterol lowers the chance of developing heart disease; hence, consuming low-LDL foods like foxtail millet is essential for ensuring adequate nutrition (Chow 2008).

Nutritional stability relies on a steady intake of essential amino acids. Lysine is a necessary amino acid for humans and animals (Tomé and Bos 2007). Therefore, the latent genetic variety of foxtail millet, which has a wide range of lysine content, may be helpful for human nutrition and animal feed. Thiamine, or vitamin B1, is essential for proper energy metabolism (Bettendorff et al. 2014). However, in economically developed countries, thiamine deficiency, which can lead to the potentially fatal disease beriberi, is uncommon due to dietary diversification and thiamine fortification of grains (Dwyer et al. 2015; Nathoo et al. 2005). However, in locations where nutritional sources of thiamine are restricted, such as Southeast Asia, thiamine deficiency does occur (Coats et al. 2012; Khounnorath et al. 2011). There is significant variation in the amount of thiamine found in foxtail millet germplasm, which suggests that it may have potential application in breeding. Increased popularity and production of foxtail millet, considered suitable food for diabetics, could contribute to reducing the worldwide prevalence of diabetes (Kam et al. 2016). In contrast to other cereal foods like rice and wheat, foxtail millet provides a steady stream of glucose without disrupting the body's metabolism (Jali et al. 2012). It is documented that incorporating foxtail millet into the diet has an antihypertensive effect (Hou et al. 2018). According to clinical trials, consuming 50 g of foxtail millet daily can reduce blood pressure, body mass index, body fat percentage, and fat mass. Additionally, this improved blood glucose levels.

3 Genetic Resources of Foxtail Millet

About 125 species make up the genus *Setaria*, which is found all over the world in warm and temperate climates (Dwivedi et al. 2012). Various phenotypic features within and across species, as well as a wide range of life cycles, ploidy levels, and breeding strategies at the genus level, suggest a highly complicated taxonomy for the *Setaria* genus, which includes both cultivated and wild species. Three separate gene

pools for the *Setaria* species have been found due to variations in genome structure. The AA genome, annotated with the genetic parameters 2n = 2x = 18, represents the primary gene pool of foxtail millet (Benabdelmouna et al. 2001a). Probably evolved from a natural cross between the diploid species *S. viridis* (green foxtail) and *S. adhaerens* (bristly grass), the weedy tetraploid species *S. faberi* (giant foxtail) and *S. verticillata* (bristly grass) both carry an AABB genome (Benabdelmouna et al. 2001b). The remaining landraces of *Setaria*, including *grisebachii, queenslandica, pumila,* and *pallide-fusca*, make up the tertiary gene pool. A diploid species known as *S. grisebachii* with a CC genome is also reported from Mexico (Wang et al. 2009). While *S. pumila* (yellow foxtail) and *S. pallide-fusca* are likewise polyploidy *Setaria* species, they lack the AA genome and are therefore not autotetraploid like *S. queenslandica* (Benabdelmouna et al. 2001a, b; Benabdelmouna and Darmency 2003). However, *S. pumila* and *S. italica* have been recognized as separate species of *Setaria* based on the appearance of their inflorescence. Both the *viridis* and the *italica* subspecies belong to the species *S. italica*. Once again, the *italica* subspecies have been classified into three races and ten subraces (Upadhyaya et al. 2011).

4 Molecular Marker Studies and QTLs Associated with Nutritional Value

The development of DNA-based molecular markers has aided in mapping genes and quantitative trait loci (QTLs) for use in plant breeding and producing genetic maps. In foxtail millet, substantial work has been conducted on developing large-scale genome-wide markers, generating maps, and tagging agronomically important genes and QTLs (Muthamilarasan et al. 2015). Intron-length polymorphisms, transposable-elements-based markers, microRNA (miRNA)-based markers, and expressed sequence tag (EST)-derived simple sequence repeats (SSRs) have all been developed due to the availability of the foxtail millet genome sequence (Kumari et al. 2013; Muthamilarasan et al. 2014; Pandey et al. 2013; Yadav et al. 2014; Zhang et al. 2014). A gene that is responsible for the "spikelet-tipped bristles" (*stb*) trait in foxtail millet was mapped by Sato et al. in 2013. This trait is essential in determining grain yield (grain number per panicle) in foxtail millet. By utilizing two F_2 populations, as well as transposon display (TD) markers and SSR markers, the location of the *stb1* gene on chromosome 2 was determined (Sato et al. 2013). Using information from the foxtail millet genome sequence, they also produced unique SSR markers, formed nine linkage groups with a total length of 1287.5 cM, and mapped *stb1* more precisely on chromosome 2. Earlier, Fang et al. (2016), by employing a Longgu7 x Yugu1 F_2 intraspecific population, detected 29 QTLs for 11 agronomic traits. In order to create a high-density genetic map, 167 members of a Yugu1 Longgu7 F_2 population were genotyped using 1013 SSR markers demonstrating polymorphism between Yugu1 and Longgu7 (Fang et al. 2016). There were 1035 loci on the genetic map, which covered a distance of 1318.8 cM at an average marker separation of 1.27 cM. In a natural population of 184 foxtail millet accessions from various regions, Gupta et al. (2014) found eight SSR markers on distinct chromosomes

that showed significant relationships with nine agronomic traits (Gupta et al. 2014). In order to investigate, extract, retain, and examine the untrodden genetic diversity of foxtail millet at the molecular level and discover the variability of its nutritional traits, Trivedi et al. (2018) analyzed 30 accessions from the Central Himalayan Region that possessed unique characteristics of agronomic significance. There was a wide variety in the nutritional aspects of husked grains, including dietary fiber, carbohydrate, protein, lysine, and thiamine content (Trivedi et al. 2018). The assembled germplasm repository might be used to generate nutritionally rich and agronomically favorable varieties of foxtail millet and to build strategies for harnessing utilizing unexploited genetic variety for food and nutrition security in agro-ecological regions. There have been many reports of molecular marker research on foxtail millet, but only a small number of these have focused on the nutritional characteristics of the crop.

5 Genomics-Aided Breeding for Nutritional Traits

5.1 Functional Genomic Studies

During the past decade, numerous studies and attempts have been made by the scientific community at large using various molecular genetic tools to investigate and enhance crop plants' nutrient use efficiency. Among these are the implementation of genome-wide association studies (GWAS), functional genomic approach, molecular marker-assisted breeding (MAB), and characterization of nutrient transporter genes. Functional genomic research on nutrient transporters is a crucial aspect of the various strategies implied. Crop plant yield, quality, and the plant's ability to overcome unfavorable conditions of low nutrient soils are all affected by nutrient use efficiency, which refers to the nutrient acquisition and utilization efficiency (Nieves-Cordones et al. 2020). The mechanism of nutrient transport in plants has been studied using foxtail millet as a model system. In addition, the recent research has opened up new possibilities for using foxtail millet as a C_4 model system to understand better and improve nutrient use efficiency (Ceasar 2022). To date, foxtail millet transporters for a variety of nutrients, including phosphate, potassium, boron, nitrogen, iron, zinc, ammonium, amino acid, and sugar, have been identified (Alagarasan et al. 2017; Ceasar et al. 2017; Liu et al. 2022; Nadeem et al. 2018; Wang et al. 2022; Yang et al. 2021a, b; Zhang et al. 2018). Uptake and redistribution of inorganic phosphate (Pi) have been linked to the phosphate transporter 1 (*PHT1*) family of candidate genes, and the expression patterns of these genes in foxtail millet have been studied (Ceasar et al. 2014; Nadeem et al. 2020). Twelve different PHT1 family members were discovered in foxtail millet (*SiPHT1;1–1;12*). Of these 12, downregulation of the *PHT1;2* genes significantly affected yeast growth, foxtail millet phenotype, and Pi transport, suggesting that this transporter is crucial for Pi uptake and export (Ceasar et al. 2017). Similarly, nitrate transporters involved in nitrate uptake in foxtail millet have been identified (Nadeem et al. 2018). Using forward and reverse genetic methods, a boron (B) transporter gene (*SiBOR1*) was

recently characterized in the cultivar Yugu 1 (Wang et al. 2022). This gene is an ortholog of the rice B transporter OsBOR1. Panicles showed prominent expression of *SiBOR1*. Overall, it appears that SiBOR1 is required for the development of the panicle to keep the grain yield in the foxtail millet stable.

Amino acid transporters (AATs) are essential for plant growth and development because they control the transmembrane transfer of amino acids. Long-distance amino acid transport, seed germination, quality development, sensitivity to pathogenic microbes, and abiotic stress are all examples (Yang et al. 2021a, b). Scientists have identified 94 AAT genes in foxtail millet, dividing them into 12 distinct subfamilies. A subclass of amino acid transporters, amino acid permeases (AAP) genes are closely associated with grain development and quality formation in numerous species. The amino acid contents of lysine, phenylalanine, leucine, and aspartic acid were found to be lower in an *Arabidopsis* AAP mutant (AtAAP6). The differential expression of SiAAP20, a homologous gene of AtAAP6, is proposed to explain the observed variation in leucine content between foxtail millet and other species. In a comparison of the two C4 plants, maize and foxtail millet, the expression of the sugar transporter, SWEET proteins, was found to be significantly higher in maize leaves (Liu et al. 2022). Higher sugar transport capacity from leaves to seeds in maize can be attributed to the higher expression of SWEET proteins in maize than in foxtail millet. It is possible that this is what keeps the carbohydrate content of foxtail millet so low compared to rice, wheat, and maize. Recent years have seen the isolation of candidate genes and gene families from foxtail millet, including resistance gene analogs, *SiNAC*, and the calcium sensor CaM gene. These have opened the door to developing functional markers for MAS and breeding for nutritional quality improvement and blast disease resistance (Puranik et al. 2011; Weng et al. 2009).

5.2 Genome-Wide Association Studies (GWAS)

Genome-wide association studies (GWAS) are essential to examine complex trait genetics. This method is preferable to interval mapping since it provides a higher resolution, uses past recombinations, allows mining a large number of alleles, and requires less time and effort. A recent report by Jaiswal et al. (2019) detailed for the first time the genetic determinants of ten nutritional elements in foxtail millet, including potassium, nickel, calcium, boron, magnesium, phosphorus, sulfur, zinc, manganese, and iron. A genome-wide association study (GWAS) was performed using 93 diverse accessions and 10,000 SNPs. Seventy-four marker-trait associations (MTAs) were connected with the ten elements listed above, with ten MTAs displaying high confidence (those associated with B, Mg, Zn, and Fe). In addition, a sizeable pyramiding effect demonstrated that connected elements might be substantially enhanced by merging numerous MTAs (Jaiswal et al. 2019). High-throughput multi-omics techniques and platforms are being developed, which will aid in metabolically deciphering the mysteries of domestication (Zhan et al. 2022). Li et al. (2022) attempted to establish a relationship between the transcriptome, metabolome, and anti-inflammatory properties of foxtail millet through a multi-omics

investigation. In order to gain a better understanding of the genetic and metabolomic variations among foxtail millet accessions, the research team conducted a high-throughput and comprehensive study on a diverse collection of foxtail millet germplasms (398 geographically diverse accessions) from China. Profiling these accessions' metabolomes revealed a significant difference in metabolites between foxtail millet and its wild relative *S. viridis*. Compared to foxtail millet, *S. viridis* displayed more fluctuation and lower overall levels of 1104 metabolites. These metabolites included serotonin, N-acetylserotonin, N-(p-coumaroyl) serotonin, trimethoprim, etc. Researchers also performed a metabolite genome-wide association study (mGWAS) to understand the genetic basis for the observed diversity in metabolic characteristics among millets. Possible causes of variance in metabolic characteristics were narrowed down to 511 genes, which were linked to 692 lead SNPs (Li et al. 2022). Li et al. (2022) used mGWAS data along with genetic and biochemical evidence to show that the foxtail millet gene *SiPSY1* encodes a limiting enzyme in the carotenoid biosynthesis pathway, thereby contributing to the millet's yellow color and the production of carotenes.

6 Traditional Practices and the Necessity of Protecting Them

Genetic resources, especially Plant Genetic Resources for Food and Agriculture (PGRFA), are essential and significant in ensuring food security. These are vital to sustainable agriculture. In addition, climate change impacts like drought, global warming, salt, temperature, and sporadic floods are the primary challenges to dependable crop production. In this way, plant genetic resources (PGR) and traditional knowledge (TK) contribute to the resilience of food crop production in the face of climate change. Despite widespread knowledge of the health benefits of millet, it is generally only consumed by indigenous peoples. This is because there are not enough convenient, accessible, or ready-to-eat alternatives to staples like rice and wheat. There have been recent efforts to make millets more accessible to consumers in more accessible forms, primarily due to their high fiber content (Deshpande and Poshadri 2011). Bread (fermented or unfermented), porridges, and snack foods are just some of the many traditional foods that can be made from millet in many parts of Africa and Asia, especially among the less privileged members of those regions' societies (Chandrasekara et al. 2012).

As a precaution against drought during sowing, Chinese farmers traditionally keep a variety of landraces that mature at different rates or are resistant to heat (Li and Wu 1996). In addition, a traditional practice of keeping specific cultivars with a short life span is also practiced in China as a precaution against drought and other natural disasters like frost and pests. Since plant thinning by hand is so time-consuming and labor-intensive, using landraces with varying growth rates could facilitate more efficient scheduling of sowing, management, and harvest. Over many years of farming, farmers developed methods for choosing the best seeds for the following year's planting. Farmers choose the best panicles in the field according to their favored ideotypes (Li and Wu 1996). The chosen panicles were allowed to dry in the open air.

As a result of the spikelets in the middle of the panicle beginning to blossom and produce grains earlier than the rest, these spikelets can obtain more nutrients that have been translocated from other organs. Each end of the panicle was removed, and the remaining parts were combined and threshed. Farmers also placed a premium on seed purity, focusing extensively on eliminating genetically diverse seed samples through a process known as "seed picking." Sun-drying, drought hardening (soaking and drying two or three times), and fumigation by burning toxic plants were common pre-sowing seed treatments used in China to improve germination vigor, drought resistance, and disease prevention, respectively (Li and Wu 1996).

Foxtail millet is also known as *tarreang* in Indonesia. This region is home to a number of local cultivars of foxtail millet, many of which are cultivated and processed by the local people into a wide range of traditional foods (Ramlah and Daryono 2020). During the festival season, the foxtail millet porridge known as *ule-uleq* is usually prepared. In addition to being processed into porridge, *tarreang* can also be processed into various other forms of local food that are processed using coconut milk. Some examples of these forms include *Sokkoltarreang*, *Jelly tarreang*, *Buras tarreang*, *Jepagollamamea*, *Jepaanjoroi*, and *Dodoltarreang* (Ramlah and Daryono 2020). Foxtail millet is the primary ingredient in Nigeria's national dish, *tuwanaduwa*, which is prepared using traditional methods (Issoufou et al. 2017). *Korramurukulu* is the name given to a beverage made from foxtail millet that does not contain any alcohol (Ramashia et al. 2021). Foxtail millet and Bengal gram flour are the two primary ingredients in the preparation of *korramurukulu*.

PGR and TK are being utilized commercially by stakeholders in developed nations due to the rapid development of technology in various domains (Salgotra and Gupta 2016). Both developing and developed countries can benefit from indigenous peoples' wisdom. However, the rewards accruing from TK use are not being distributed fairly. Developed country's stakeholders have granted IPRs on the vast majority of PGR and related sources, with benefits being shared with local and indigenous communities. However, most nations' intellectual property laws are not robust enough to safeguard their PGR and TK. The debate over IP protection for the TK of indigenous farming communities has opened a new line of inquiry into moral concerns. Due to the unawareness of TK and IP laws protecting them, most of these issues' international disputes remained contentious and unresolved. It is imperative that these issues be discussed at the international level with world organizations, stakeholders, and NGOs to find solutions to these complex problems. Indigenous peoples should be provided with up-to-date information through IPR-related awareness programs (Salgotra and Gupta 2016).

7 Conclusion and Future Prospects

Foxtail millet, one of the world's oldest and most widely grown crops, has a high tolerance for soils with low fertility. It serves as an excellent experimental platform for researching the genetics of C_4 plants. Functional genomic investigations were

hastened by the availability of the genome sequence, leading to the characterization of various dietary features. Foxtail millet's anti-diabetic, antihypertensive, anti-hyperlipidemic, and antioxidative potential makes it a vital part of any healthy diet in a society where lifestyle disorders are so prevalent. Foxtail millet, formerly an overlooked and understudied crop, is now recognized as a high-quality genetic resource with many practical applications. Although various functional genomics, proteomics, metabolomics, and population studies have been reported on foxtail millet, the number of studies concentrating on the millet's nutritional properties is minimal. Applying omics or multi-omics techniques to enhance the nutritional quality of foxtail millet could potentially open up new avenues for plant breeding initiatives using this ancient grass crop. The substantial diversity of molecular clustering and nutritional features and the correlation between the two can be utilized to build further techniques tailored to produce nutritionally dense variations.

Acknowledgments Authors acknowledge financial support from Council of Scientific & Industrial Research (CSIR) with grant No. NIScPR/OLP/0030/2022. Acknowledgment is due to the director, CSIR-NIScPR, New Delhi, for providing support for the study.

References

Akoh CC, Min DB (2007) Food lipids: chemistry, nutrition and biotechnology, 3rd edn. Marcel Dekker, New York

Alagarasan G, Dubey M, Aswathy KS, Chandel G (2017) Genome wide identification of orthologous ZIP genes associated with zinc and iron translocation in Setaria italica. Frontiers. Plant Sci 8:775

Bandyopadhyay T, Jaiswal V, Prasad M (2017) Nutrition potential of foxtail millet in comparison to other millets and major cereals, pp 123–135

Benabdelmouna A, Abirached-Darmency M, Darmency H (2001b) Phylogenetic and genomic relationships in Setaria italica and its close relatives based on the molecular diversity and chromosomal organization of 5S and 18S-5.8S-25S rDNA genes. Theor Appl Genet 103: 668–677

Benabdelmouna A, Darmency H (2003) Copia-like retrotransposons in the genus *Setaria*: Sequence heterogeneity, species distribution and chromosomal organization. Plant Syst Evol 237:127–136

Benabdelmouna A, Shi Y, Abirached-Darmency M, Darmency H (2001a) Genomic in situ hybridization (GISH) discriminates between the A and the B genomes in diploid and tetraploid *Setaria* species. Genome 44:685–690

Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J et al (2012) Reference genome sequence of the model plant Setaria. Nat Biotechnol 30(6):555–561

Bettendorff L, Lakaye B, Kohn G, Wins P (2014) Thiamine triphosphate: a ubiquitous molecule in search of a physiological role. Metab Brain Dis 29(4):1069–1082

Ceasar SA (2022) Foxtail millet (Setariaitalica) as a model system to study and improve the nutrient transport in cereals. Plant Growth Regul 1:1–8

Ceasar SA, Baker A, Ignacimuthu S (2017) Functional characterization of the PHT1 family transporters of foxtail millet with development of a novel agrobacterium-mediated transformation procedure. Sci Rep 7(1):14064

Ceasar SA, Hodge A, Baker A, Baldwin SA (2014) Phosphate concentration and arbuscular mycorrhizal colonisation influence the growth, yield and expression of twelve PHT1 family phosphate transporters in foxtail millet (*Setaria italica*). PLoS One 9(9):e108459

Chandel G, Meena RK, Dubey M, Kumar M (2014) Nutritional properties of minor millets: neglected cereals with potentials to combat malnutrition. Curr Sci 107(7):1109–1111

Chandrasekara A, Naczk M, Shahidi F (2012) Effect of processing on the antioxidant activity of millet grains. Food Chem 133(1):1–9

Chow CK (ed) (2008) Fatty acids in foods and their health implications, 3rd edn. CRC Press. Coats, Boca Raton, FL

Coats D, Shelton-Dodge K, Ou K, Khun V, Seab S et al (2012) Thiamine deficiency in Cambodian infants with and without beriberi. J Pediatr 161(5):843–847

Deshpande HW, Poshadri A (2011) Physical and sensory characteristics of extruded snacks prepared from foxtail millet based composite flours. Int Food Res J 18:751–756

Dwivedi S, Upadhyaya H, Senthilvel S, Hash C, Fukunaga K et al (2012) Millets: genetic and genomic resources. In: Janick J (ed) Plant breed rev, vol 35. John Wiley & Sons, Inc., USA, pp 247–375

Dwyer JT, Wiemer KL, Dary O, Keen CL, King JC et al (2015) Fortification and health: challenges and opportunities. Adv Nutr. (Bethesda, Md 6(1):124–131

Eastwood M (2003) Principles of human nutrition, 2nd edn. Blackwell Science, Malden, MA

Fang X, Dong K, Wang X, Liu T, He J et al (2016) A high density genetic map and QTL for agronomic and yield traits in foxtail millet [Setariaitalica (L.) P. Beauv.]. BMC Genom 17(1):1–12

Favalli EG (2020) Understanding the role of Interleukin-6 (IL-6) in the joint and beyond: a comprehensive review of IL-6 inhibition for the Management of Rheumatoid Arthritis. Rheumatol Ther 7(3):473–516

Flight I, Clifton P (2006) Cereal grains and legumes in the prevention of coronary heart disease and stroke: a review of the literature. Eur J Clin Nutr 60(10):1145–1159

Guo Y, Wang B, Wang T, Gao L, Yang ZJ et al (2020) Biological characteristics of il-6 and related intestinal diseases. Int J Biol Sci 17(1):204–219

Gupta S, Kumari K, Muthamilarasan M, Parida SK, Prasad M (2014) Population structure and association mapping of yield contributing agronomic traits in foxtail millet. Plant Cell Rep 33 (6):881–893

Hariprasanna K, Jacob J, Patroti P, Visarada KBRS (2017) Breeding strategies in foxtail millet. In: Prasad M (ed) The foxtail millet genome. Springer International Publishing, pp 53–61

Hou D, Chen J, Ren X, Wang C, Diao X et al (2018) A whole foxtail millet diet reduces blood pressure in subjects with mild hypertension. J Cereal Sci 84:13–19

Issoufou A, Mahamadou E, Tidjani A, Moutaleb O, Guo-Wei L (2017) Functional and sensory properties of *Tuwan Aduwa*: a traditional meal made from foxtail millet and desert date extract. Niger Food J 35(1):1–7

Jaiswal V, Bandyopadhyay T, Gahlaut V, Gupta S, Dhaka A et al (2019) Genome-wide association study (GWAS) delineates genomic loci for ten nutritional elements in foxtail millet (Setariaitalica L.). J Cereal Sci 85:48–55

Jali MV, Kamatar MY, Jali SM, Hiremath MB, Naik RK (2012) Efficacy of value added foxtail millet therapeutic food in the management of diabetes and dyslipidemia in type 2 diabetic patients. Recent Res Sci Technol 4(7):3–04

Kam J, Puranik S, Yadav R, Manwaring HR, Pierre S et al (2016) Dietary interventions for type 2 diabetes: how millet comes to help. Front Plant Sci 7:1454

Khounnorath S, Chamberlain K, Taylor AM, Soukaloun D, Mayxay M et al (2011) Clinically unapparent infantile thiamin deficiency in Vientiane, Laos. PLoS Negl Trop Dis 5(2):e969–e969

Kumari K, Muthamilarasan M, Misra G, Gupta S, Subramanian A et al (2013) Development of eSSR-markers in *Setaria italica* and their applicability in studying genetic diversity, cross-transferability and comparative mapping in millet and non-millet species. PLoS One 8(6):e67742

Lata C, Gupta S, Prasad M (2013) Foxtail millet: a model crop for genetic and genomic studies in bioenergy grasses. Crit Rev Biotechnol 33:328–343

Lata C, Shivhare R (2017) Genetic Determinants of Abiotic Stress Tolerance in Foxtail Millet. In: Prasad, M. (eds) The Foxtail Millet Genome. Compendium of Plant Genomes. Springer, Cham. 85–104

Li S, Dong X, Fan G, Yang Q, Shi J et al (2018) Comprehensive profiling and inheritance patterns of metabolites in foxtail millet. Front Plant Sci 871:1716

Li X, Gao J, Song J, Guo K, Hou S et al (2022) Multi-omics analyses of 398 foxtail millet accessions reveal genomic regions associated with domestication, metabolite traits, and anti-inflammatory effects. Mol Plant 15(8):1367–1383

Li Y, Wu S (1996) Traditional maintenance and multiplication of foxtail millet (*Setaria italica* (L.) P. Beauv.) landraces in China. Euphytica 87(1):33–38

Liu Z, Fan H, Ma Z (2022) Comparison of SWEET gene family between maize and foxtail millet through genomic, transcriptomic, and proteomic analyses. The Plant Genome 15(3):e20226

Muthamilarasan M, Dhaka A, Yadav R, Prasad M (2015) Exploration of millet models for developing nutrient rich graminaceous crops. Plant Sci 242:89–97

Muthamilarasan M, Venkata Suresh B, Pandey G, Kumari K, Parida SK, Prasad M (2014) Development of 5123 intron-length polymorphic markers for large-scale genotyping applications in foxtail millet. DNA Res 21(1):41–52

Nadeem F, Ahmad Z, Ul Hassan M, Wang R, Diao X et al (2020) Adaptation of foxtail millet (*Setaria italica* L.) to abiotic stresses: a special perspective of responses to nitrogen and phosphate limitations. Frontiers. Plant Sci 11:187

Nadeem F, Ahmad Z, Wang R, Han J, Shen Q et al (2018) Foxtail millet [*Setaria italica* (L.) beauv.] grown under low nitrogen shows a smaller root system, enhanced biomass accumulation, and nitrate transporter expression. Frontiers. Plant Sci 9:205

Nathoo T, Holmes CP, Ostry A (2005) An analysis of the development of Canadian food fortification policies: the case of vitamin B. Health Promot Int 20(4):375–382

Navarro-González JF, Mora-Fernández C (2008) The role of inflammatory cytokines in diabetic nephropathy. J Am Soc Nephrol 19(3):433–442

Nieves-Cordones M, Rubio F, Santa-María GE (2020) Editorial: nutrient use-efficiency in plants: an integrative approach. Front Plant Sci 11:623976

Pandey G, Misra G, Kumari K, Gupta S, Parida SK et al (2013) Genome-wide development and use of microsatellite markers for large-scale genotyping applications in foxtail millet [Setaria italica (L.)]. DNA Res 20(2):197–207

Puranik S, Bahadur RP, Srivastava PS, Prasad M (2011) Molecular cloning and characterization of a membrane associated NAC family gene, SiNAC from foxtail millet [*Setaria italica* (L.) P. Beauv]. Mol Biotechnol 49(2):138–150

Ramashia SE, Mashau ME, Onipe OO (2021) Millets cereal grains: nutritional composition and utilisation in sub-Saharan Africa. Cereal Grains 1–14

Ramlah PMB, Daryono BS (2020) Local food diversification of foxtail millet (*Setaria italica*) cultivars in West Sulawesi, Indonesia: a case study of diversity and local culture. Biodiversitas 21(1):67–73

Roumani M, Duval RE, Ropars A, Risler A, Robin C et al (2020) Phenolamides: plant specialized metabolites with a wide range of promising pharmacological and health-promoting interests. Biomed Pharmacother 131:110762

Saleh ASM, Zhang Q, Chen J et al (2013) Millet grains: nutritional quality, processing, and potential health benefits. Compr Rev Food Sci Food Saf 12:281–295

Salgotra RK, Gupta BB (2016) Plant genetic resources and traditional knowledge for food security. In: Plant genetic resources and traditional knowledge for food security. Springer, Singapore

Sato K, Mukainari Y, Naito K, Fukunaga K (2013) Construction of a foxtail millet linkage map and mapping of spikelet-tipped bristles 1(stb1) by using transposon display markers and simple sequence repeat markers with genome sequence information. Mol Breed 31(3):675–684

Singh P, Raghuvanshi RS (2012) Finger millet for food and nutritional security. Afr J Food Sci 6(4): 77–84

Tomé D, Bos C (2007) Lysine requirement through the human life cycle. J Nutr 137(6 Suppl 2): 1642S

Trivedi AK, Arya L, Verma SK, Tyagi RK, Hemantaranjan A et al (2018) Molecular profiling of foxtail millet (Setariaitalica (L.) P. Beauv) from central Himalayan region for genetic variability and nutritional quality. J Agric Sci 156(3):333–341

Ullah A, Munir S, Badshah SL, Khan N, Ghani L et al (2020) Important flavonoids and their role as a therapeutic agent. Molecules 25(22):5243

Upadhyaya HD, Ravishankar CR, Narasimhudu Y, Sarma NDRK, Singh SK et al (2011) Identification of trait-specific germplasm and developing a mini core collection for efficient use of foxtail millet genetic resources in crop improvement. Field Crop Res 124:459–467

Wang H, Tang S, Zhi H, Xing L, Zhang H et al (2022) The boron transporter SiBOR1 functions in cell wall integrity, cellular homeostasis, and panicle development in foxtail millet. Crop J 10(2):342–353

Wang J, Zhang G, Liu X, Quan Z, Cheng S et al (2012) Genome sequence of foxtail millet (Setariaitalica) provides insights into grass evolution and biofuel potential. Nat Biotechnol 30 (6):549–554

Wang Y, Zhi H, Li W, Li H, Wang Y et al (2009) A novel genome of C and the first autotetraploid species in the *Setaria* genus identified by genomic in situ hybridization. Genet Resour Crop Evol 56(6):843–850

Weng Q, Li Z, Xing J, Dong Z, Dong J (2009) Cloning and structure analysis of an NBS-LRR disease-resistant gene from Setaria italica Beauv. Front Agric China 3(3):240–246

Yadav CB, Muthamilarasan M, Pandey G, Khan Y, Prasad M (2014) Development of novel microRNA-based genetic markers in foxtail millet for genotyping applications in related grass species. Mol Breed 34(4):2219–2224

Yang Y, Chai Y, Liu J, Zheng J, Zhao Z et al (2021b) Amino acid transporter (AAT) gene family in foxtail millet (Setariaitalica L.): widespread family expansion, functional differentiation, roles in quality formation and response to abiotic stresses. BMC Genomics 22(1):519

Yang L, Li R, Cui Y, Qin X, Li Z (2021a) Comparison of nutritional compositions of foxtail millet from the different cultivation regions by UPLC-Q-Orbitrap HRMS based metabolomics approach. J Food Biochem 45(10):e13940

Yousaf L, Hou D, Liaqat H, Shen Q (2021) Millet: a review of its nutritional and functional changes during processing. Food Res Int 142(17):110197

Zhan C, Shen S, Yang C, Liu Z, Fernie AR et al (2022) Plant metabolic gene clusters in the multi-omics era. Trends Plant Sci 0(0):981

Zhang S, Tang C, Zhao Q, Li J, Yang L et al (2014) Development of highly polymorphic simple sequence repeat markers using genome-wide microsatellite variant analysis in foxtail millet [Setariaitalica (L.) P. Beauv.]. BMC Genomics 15(1):1–9

Zhang H, Xiao W, Yu W, Yao L, Li L et al (2018) Foxtail millet SiHAK1 excites extreme high-affinity K+ uptake to maintain K+ homeostasis under low K+ or salt stress. Plant Cell Rep 37 (11):1533–1546

Genetic and Genomic Resources for Harnessing the Health-Related Genes in Finger Millet

S. Antony Ceasar and B. Kalyan Babu

Contents

Abstract

Finger millet (*Eleusine coracana* (L.) Gaertn) is a nutrient-dense cereal crop grown and consumed by resource-poor farmers in India. Finger millet grains contain a higher content of minerals such as calcium (Ca), phosphorus (P), iron (Fe), and manganese (Mn) compared to other major cereals. Notably, it has a tenfold higher Ca in seeds than other major cereals. This chapter covers the wider information on finger millet, including nutrient profile, nutritional importance, genetic resources of health-related (HR) genes, genetic diversity in HR gene-rich germplasm, molecular mapping of HR genes, genomics-aided breeding for HR traits, present concepts and strategies developed including genetic engineering

S. A. Ceasar (✉)
Division of Plant Molecular Biology and Biotechnology, Department of Biosciences, Rajagiri College of Social Sciences, Kerala, Kochi, India
e-mail: saceasar@rajagiri.edu

B. K. Babu
ICAR-Indian Institute of Oil Palm Research, Pedavegi, Andhra Pradesh, India

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_9

and genome editing by clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein (Cas) system, and role of bioinformatic tools for finger millet improvement. Especially the genome sequence of finger millet was released recently, and the completely annotated genome sequence is also freely available for researchers for mining gene and protein sequences. The completely annotated genome sequence will aid in high-resolution genetic studies. The annotated genome sequence will also help apply the CRISPR/Cas tool for the functional characterization of key genes, such as those involved in grain calcium accumulation in finger millet. Thus, it will help improve calcium transport in other millets and non-millet cereals in the future.

Keywords

Nutrition · Health-related genes · Molecular breeding · Genetic engineering · Genome editing

1 Introduction

The growing world population has put pressure on the food demand. The Food and Agriculture Organization (FAO) estimates that the world population may increase by 10 billion in 2050 (FAO 2017). Food demand is expected to increase between 59% and 98% by 2050 (McKenzie and Williams 2015). We need to improve food production to meet the food demand, especially amid the adverse effects of climate change. It is not only the quality of the crop production but also the quantity of the seeds that should be improved to meet the nutritional requirements. The quality of the crop is directly linked to human health (Cakmak and Hoffland 2012; Zhao et al. 2020). Worldwide, over three billion people suffer from various health problems due to low nutrient content in their diet (Krishna et al. 2022a, b). The major cause of malnutrition problems is the low nutrient content in the consumed food grains, mainly affecting children and pregnant women (WHO 2016). Enrichment of nutrient content in the edible part of the crop is a major area of research in the present scenario. Therefore, agriculture scientists are trying to improve crop quality in different ways.

The finger millet (*Eleusine coracana* [L.] Gaertn.) is one of the most important foods and nutritional security crops of the future (Ceasar et al. 2018; Maharajan et al. 2021, 2022a). It is well recognized worldwide due to its nutritional properties and related health benefits (Cakmak and Hoffland 2012; Thapliyal and Singh 2015). It has higher nutritional properties than other major staple grains such as rice and wheat (Shobana et al. 2013). Finger millet contains a rich source of calcium (Ca) (0.38%), other minerals (2.7%), dietary fiber (18%), phenolic compounds (0.3–3%), etc. (Poonia et al. 2012; Shobana et al. 2013; Thapliyal and Singh 2015). They are popularized for human health benefits such as antidiabetic, anti-tumorigenic, atherosclerogenic, antioxidant, and antimicrobial properties (Gull et al. 2014). Due to their nutritional value, finger millet is called nutria-millet or nutria-cereals (Swaminaidu et al. 2015). Further improvement of finger millet may provide food and nutritional security on a long-term basis in the future.

Plant breeding is essential for crop improvement. In the past, phenotype-based (conventional) breeding approaches significantly impacted crop improvement (Patra et al. 2020). Usually, it is a time-consuming process, and phenotype-based breeding approaches hamper the efficiency of crop improvement due to environmental factors (Krishna et al. 2021). Advanced finger millet research may help identify quantitative trait loci (QTLs)/genes/alleles related to nutritional traits. It provides the opportunity to improve the finger millet through breeding approaches. Nowadays, next-generation sequencing (NGS) technology provides the opportunity for precision breading through genome-assisted breeding (GAB) (Van et al. 2013). The NGS platform helps to understand the genome organization of crops through genotyping-by-sequencing (GBS) technology (Krishna et al. 2020). It easily identifies desirable traits on a genetic basis and improves breeding efficiency. It is considered a next-generation breeding approach (Ray and Satya 2014). We hope that the genome-based breeding approaches help improve finger millet quality and strengthen nutritional security.

A complete and fully annotated genome sequence is also available for finger millet which provides an opportunity to design advanced forward and reverse genetic studies like those performed in other model cereals like rice. It also helps to mine genes and proteins for any novel work. The genome sequence is also helpful in applying genome editing tools like clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein (Cas) system in finger millet for targeting specific genes. Genome editing will help characterize the key genes in finger millet and transfer such genes to other millets and cereals.

2 Nutritional Importance

With the emergence of more diabetic and coronary diseases among the world population, healthy food is becoming prime important worldwide. During the old days, people consumed many types of millet as part of their food, along with unpolished rice and wheat, which has a major role in maintaining a healthy life. However, with the advancement of technology and busy life, people are crazier toward highly polished rice, not considering millets as part of their food since there is a misconception that those are food for poor people only. Because of this, many health issues are emerging which can be tacked eco-friendly through nutritious crops. In this connection, nowadays, more awareness is created among people to consume millets, especially finger millet, for reducing diabetes, coronary health issues, and obesity and to maintain a good healthy life. In the present "nutrigenomics" era, integrating nutrition values with genomics has become popular for gene identification, cloning, and marker-assisted selection. Finger millet is considered a nutrition hub of many nutritional parameters like a high amount of Ca and protein and rich in fiber, aiding in reducing malnutrition in many developing and underdeveloped nations (Millet network of India 2010). National academies of the United States also considered finger millet as "super cereal," the most nutritious cereal crop (National Research Council 1996). It is known for its nutritional value mainly because of its rich source of Ca, which is ten times more than wheat, maize, or brown rice (Antony and Chandra 1998) and three times more Ca in

milk. Ca helps in vascular and muscular strengthening, reducing certain types of cancer, and reducing the weight of the body and also helps in the reduction of diabetes (Pittas et al. 2006). On average, finger millet consists of 340–360 mg/100 g Ca content (Shobana et al. 2013). Along with Ca, it has a high amount of iron (Fe), polyphenols like phytochemicals, and polyunsaturated fatty acids. It is also loaded with high essential amino acids like tryptophan, lysine, and methionine (McDonough et al. 2000) which are very much required for human growth. These essential amino acids are not produced by human metabolism and hence should be supplemented by external sources like food supplements like finger millet. Recently, it is called "super cereal" by some authors (Kumar et al. 2016) due to its high nutritional properties.

Human diseases like diabetes, chronic coronary cardiovascular diseases, and cancer are closely associated with reactive oxygen and nitrogen species. The seed coat in the grain of finger millet consisted of high phenolics like benzoic acid having antioxidant properties, as reported by Chandrasekara and Shahidi (2010). Even among finger millets, brown varieties have more phenolic compounds than the white seed coat finger millet varieties Subba Rao and Muralikrishna (2002). Finger millet can help in bringing down aging (Hegde et al. 2002) by shrinking the toughness of the tendon and blood vessel tissues. Finger millet also has ferulic acid and vitamin B17, which are reported to help treat cancer (Kawabata et al. 2000) and in the case of esophageal cancer (Griffin 1974). The abundance of calcium and magnesium (Mg) in finger millet decreases the type 2 diabetes risk by reducing the glucose levels in the blood. Even consuming a lesser amount of methanolic extracts of nearly 3 mg from finger millet flour aids insignificantly inhibiting glycation, which is a major part of diabetes pathogenesis (Hegde et al. 2002).

People consuming millets, especially finger millet, as part of their routine food have fewer chances for cardiovascular disease, which is proved in many reports worldwide (Lee et al. 2010). The fiber component of the finger millet reduces the low-density lipoprotein (LDL) cholesterol (bad cholesterol) and triglycerides, which are the main causes of heart-related problems. These soluble and insoluble fibers of finger millet seed coat and grains also improve digestibility and reduce gastrointestinal problems, which has become a substitute for wheat and rice since it is gluten-free (Chandrasekara and Shahidi 2012). It also arrests constipation due to its high cellulose content. Because of its multi-nutritional advantages like high Ca, protein, fiber content, antioxidant properties, high vitamins, and essential amino acids, it is popularly called super cereal. A brief review of the nutritional properties of whole millet is available (Kumar et al. 2016), which readers can access to enhance their knowledge.

3 Genetic Resources of Health-Related (HR) Genes

Finger millet has rich genetic material like landraces, wild cultivars, and traditional varieties grown by farmers who are not fully exploited for their effective use in finger millet breeding programs. All these genetic materials have very good nutritional importance, climate-resistant alleles, high genetic variability, and other useful

parameters. However, compared to major cereals like rice, maize, and wheat, the germplasm of finger millet is not much used to that extent due to less research focus on millet crops, especially finger millet, where few genetic studies were carried out.

At the global level, only one organization, i.e., International Crops Research Institute (ICRISAT), Hyderabad, a part of the Consultative Group on International Agricultural Research (CGIAR) organizations, has a good amount of genetic material collection of approximately 6000 accessions (Sood et al. 2016). Upadhyaya et al. (2006) conducted an exhaustive study to sort out this whole germplasm into a small core collection representing 6000 accessions regarding genetic diversity and morphological characteristics like qualitative and quantitative parameters. This core collection has nearly 622 accessions. Further, Upadhyay et al. (2011) developed a mini-core collection to reduce the number of accessions for easy analysis studies while maintaining all the essential parameters like genetic variation, quality parameters, and quantitative traits. The genetic material in the collection originates from South America, Asia, Europe, and Africa. The mini-core has germplasm with the origin of countries like Uganda, India, Tanzania, and Kenya. These core and mini-core collections are being shared with many research groups in the nation's interest for various studies on material transfer agreements. Babu et al. (2014) identified more than 100 genotypes having more than 0.85% of tryptophan, which can be used as a good source of essential amino acids. It was reported that genotypes like GE1437, IE6240, and GE2136 had a high amount of tryptophan and lysine and would probably be the best source to promote healthy finger millet germplasm. Likewise, they also found GPHCPB1, GE1680, and GE1621 with more protein content. The above-reported germplasm can be effectively used for alleviating malnutrition in the country. Germplasm with special traits can be used for many breeding purposes to develop abiotic and biotic stress-resistant hybrids with good nutritional properties. Yadav et al. (2020) reported very good germplasm with a high amount of calcium content, viz., GPHCPB45, GPHCPB44, and GPHCPB31, which can be used directly for avoiding malnutrition in developing and undernourished countries. This type of genetic material can be the best source of nutritional food in the world for reducing health-related issues.

4 Genetic Diversity in HR Gene-Rich Germplasm

The beginning of molecular studies in finger millet started in 1995 for studying the molecular divergence and origin of five finger millet species using first-generation molecular markers like random amplified polymorphic DNA (RAPD) and inter-simple sequence repeat (ISSR) markers (Salimath et al. 1995). However, the finger millet genetic diversity pace went very slow compared to major cereals like maize, rice, and wheat. Later in 2004, the assessment of diversity at the molecular level started again using RAPD and ISSR markers in India (Fakrudin et al. 2004). From that time onward, several workers started exploiting the landraces and locally available germplasm for molecular studies (Babu et al. 2007; Dida et al. 2008).

All these genetic diversity analyses were for morphological traits, mostly for their pre-breeding programs. However, from 2010 onward, there has been a focus on assessing molecular diversity for nutritional property germplasm using various molecular markers like RAPD, ISSR, expressed sequence tag-derived simple sequence repeat (EST-SSR), and simple sequence repeat (SSR) markers. Panwar et al. (2010), for the first time, used RAPD, SSR, and cytochrome P450 markers for the genetic diversity of 52 accessions sourced from the northwestern Himalayan region of India. The clustering pattern differentiated low-, medium-, and high-calcium-content genotypes using these markers, and they observed a very wide genetic base among the germplasm. Nirgude et al. (2014) observed high variation in protein content, and using EST-SSR markers, they studied the genetic level variation among finger millet germplasm. This study reported molecular markers for genetic variations for grain protein and Ca contents. For the Ca content of finger millet grains, several reports are available which are focused on using mainly SSR and single nucleotide polymorphism (SNP) markers (Sharma et al. 2017; Yadav et al. 2020). Sharma et al. (2017) studied the population structure of finger millet accessions collected from different regions of India, including mini-core collection. Babu et al. (2014) studied the genetic diversity and population structure of a large number of germplasm (190) using candidate gene-based EST markers and genomic markers for tryptophan and lysine content. These essential amino acids are required for human health as described in the earlier section, i.e., nutrition importance.

5 Genetics and Molecular Mapping of HR Genes

As described earlier, finger millet is a super cereal with many nutritional properties. Gene/QTL identification for these nutritional properties plays an important role in marker-assisted breeding programs. QTL mapping was started in 1998 by the John Innes Centre, a research group using restriction fragment length polymorphism (RFLP) markers, covering a total of 1100 cM distance into 27 linkage groups (Gale and Devos 1998). After that, finger millet research took a long time to construct other genetic maps. Genetic maps play a crucial role in any breeding and marker-assisted breeding for the generation of improved varieties with desirable traits. From 2007 onward, there were no genetic mapping studies for QTL identification until 2011. Bharathi (2011) studied a large number of germplasm for QTL identification on morphological traits using a limited number of SSR markers. Since the finger millet genome was not sequenced till 2018, most of the focus of the researchers was to exploit the sequence information available in closely related crops like rice, maize, wheat, and foxtail millet. For nutrition-related traits, Babu et al. (2014) found significant loci/QTL for protein and tryptophan content for the first time. They found one EST base marker (OM5) linked to tryptophan content related to the 27-KDa *gamazein* gene of maize. Likewise, the results detected one more significant QTL for the RISBZ candidate gene linked by FMO2EST1 which explained 9% of phenotypic variance. Kumar et al. (2016) identified SNP markers

linked to candidate genes of Ca transporter for variation in whole-grain Ca content. They resolved 23,000 SNPs from 33 GB of sequenced information. Sharma et al. (2017) reviewed calcium nutrition value in finger millet, its bioavailability, and genomics. Yadav et al. (2020) identified significant loci influencing Ca content through association mapping. Their results showed that UGEP60 could be a potential molecular marker to detect high Ca content lines in finger millet.

6 Genomics-Aided Breeding for HR Traits

Humans mainly depend on plant-based food for their daily lives (Sands et al. 2009). Good-quality plant foods are essential to maintain good health (Welch 2002). It also leads to preventing malnutrition issues worldwide. Plant breeders are working to develop new varieties of crops that are beneficial to human health through innovative breeding approaches. In the past, conventional breeding approaches have contributed to improving the quality of many crop plants (Tahir ul Qamar et al. 2020). But it has some drawbacks; the phenotypic-based breading is influenced by environmental factors and reduces the efficiency of crop improvement (Krishna et al. 2021). Therefore, understanding the structural and functional characteristics of the finger millet is crucial for their improvement through GAB (Krishna et al. 2022a, b). Remarkably, marker-assisted selection provided the opportunity for precise crop breeding through GAB. The genome-wide assessments of the genetic variation for the desirable traits of finger millet germplasm may help improve the quality of finger millet (Krishna et al. 2021).

NGS technology is employed for whole-genome sequencing (WGS), which helps understand the genome organization of crop plants. It helps accurately predict the genetic basis of phenotypic variation of the genotypes (Maharajan et al. 2022b). The NGS technology has provided the foundation for whole-genome scan studies such as genome-wide association studies (GWAS) and genomic selection (GS). These two approaches are helpful for the identification of desirable traits for finger millet improvement through high-throughput sequence technology like the GBS method. The GWAS and GS approaches are widely adopted for their efficiency during crop improvement. In GWAS, many valuable QTLs have been identified in cereal and non-cereal crops such as rice (Bollinedi et al. 2020; Verma et al. 2022; Zhang et al. 2020), maize (Ndlovu et al. 2021; Ruanjaichon et al. 2021; Zheng et al. 2021), wheat (Muhu-Din Ahmed et al. 2020; Rathan et al. 2022; Shi et al. 2022), sorghum (Cruet-Burgos et al. 2020; Nida et al. 2021), foxtail millet (Jaiswal et al. 2019a, b; Jia et al. 2013), proso millet (Boukail et al. 2021), pearl millet (Pujar et al. 2020; Yadav et al. 2021), etc. The GWAS can also be applied to identify QTL associated with HR traits. A little effort was paid to GWAS in finger millet compared with other cereal and non-cereal crops. However, GWAS has been used to identify finger millet's nutritional and yield-related QTL. For example, Tiwari et al. (2020) identified QTL and genes linked to seed protein and grain yield in finger millet using 2977 high-throughput SNP markers through the GBS method. In silico approaches revealed that the QTL marker was linked to the candidate gene responsible for seed protein content in finger millet

(Tiwari et al. 2020). Furthermore, the candidate gene *aspartyl protease* was found to be responsible for seed protein content in finger millet (Tiwari et al. 2020). Puranik et al. (2020) analyzed the six nutritional traits (Fe, zinc (Zn), Ca, Mg, potassium (K), and sodium (Na)) using 190 finger millet accessions by GBS method and identified the marker-trait associations (MTAs). In this study, 169,365 SNPs were generated from GBS method, and out of these 418 SNPs were associated with various nutritional traits of finger millet (Puranik et al. 2020). Sharma et al. (2018) generated SNP markers by GBS using 113 finger millet accessions and identified reliable SNP markers linked to agro-morphological traits and grain yield in finger millet. These studies provide an opportunity to improve the finger millet's nutritional quality through GWAS. The GS-related study is not yet available in the finger millet. Huge germplasm is available in finger millets; it includes cultivars, breeding materials, landraces, training populations, wild relatives, etc. Therefore, plant breeders must utilize the finger millet germplasm for GWAS and GS approaches. It may help to identify the QTL markers and candidate genes related to human health, which will strengthen nutritional security in the future. Recently, a GWAS was conducted to identify the genes related to high seed Ca content (Sharma et al. 2022). They have identified SNP markers for candidate genes, including *calmodulin-binding protein (CBP*) and *CBL-interacting protein kinase 7 (CIPK7*) of foxtail millet, and these two genes (*EcCBP* and *EcCIPK7*) were highly expressed in high calcium genotypes of finger millet (Sharma et al. 2022).

7 Recent Concepts and Strategies Developed

Genome editing is a recently added toolbox for plant scientists to precisely identify and improve crop plants' genomic regions. Especially many researchers quickly adopted the CRISPR/Cas9 system for plant improvement studies. Cas9 system has offered a quick, easy, cost-effective procedure and emerged as a potential alternative to the previously used tools like mega-nucleases (MNs), Zn-finger nucleases (ZFNs), and transcription activator-like effector nucleases (TALENs) (Doudna and Charpentier 2014; Jinek et al. 2012). Based on CRISPR/Cas9 system, which majorly works for targeted gene disruption (gene knock-out), several other variants of CRISPR systems have been developed for diverse applications such as transcriptional regulation, point mutation, and correction of a small portion in DNA. These include dead Cas9 (dCas9) for transcriptional regulation (Qi et al. 2013), CRISPR base editor for inducing point mutations (Keiji et al. 2016), and CRISPR prime editor for donor-free specific DNA editing even for small deletion or insertion mutations (Anzalone et al. 2019; Hillary and Ceasar 2022). However, these tools have not yet reached the millets, including finger millet (Ceasar 2021). CRISPR/Cas9 was the earlier tool applied widely in plant science research since it cleaves double-stranded DNA straightaway. It helps for the gene knock-out studies with a simpler construct design strategy. But this tool is not yet applied to any millet, including finger millet.

Genome editing requires a complete and fully annotated genome sequence. Genome sequence helps identify a specific gene or region of the gene to be targeted by genome editing tools. For example, CRISPR/Cas9 system requires designing specific guide

RNAs (gRNAs) to target the genomic region precisely. Thus, the gRNA design demands the plant's fully annotated genome. Now the fully annotated genome sequence of finger millet is freely available for mining genes and further annotations through the phytozome website (https://phytozome-next.jgi.doe.gov/). So the target regions for guide RNAs can be designed and applied in the finger millet for knock-out studies.

Similarly, many other CRISPR tools, like the base editor and prime editor, could also be applied in finger millet for targeted genome editing. In some cases, the annotated genome sequence is also required for resequencing the genome to confirm any off-target effects during CRISPR/Cas9 knock-out. So the annotated genome could be helpful for genome resequencing finger millet following the CRISPR/Cas9 application.

Apart from a fully annotated genome sequence, another requirement is the availability of an efficient transformation system to deliver the CRISPR constructs for the targeted genome editing. Many transformation protocols have been reported for finger millet (Vetriventhan et al. 2020; Ceasar 2021). The *Agrobacterium*-mediated transformation has predominantly occurred in finger millet (Vetriventhan et al. 2020). Direct and indirect regeneration methods were employed to recover the transgenic plants after the transformation with the concerned constructs. Shoot apex explants were predominantly used for the infection and co-cultivation with the *Agrobacterium* in the genetic transformation studies (Vetriventhan et al. 123 AD). So the transformation protocols reported for the finger millet will be useful for introducing CRISPR constructs and developing the mutant finger millet. CRISPR/Cas system has been harnessed to develop plants with superior traits in the past (Song et al. 2016). This system will be very useful in characterizing key genes involved in the Ca transport to the seeds. Such studies will help identify key genes and transfer such genes to other millets and non-millet cereals for improving nutrient transport and seed fortification.

8 Brief on Genetic Engineering for HR Traits

Significant research progress has been made in genetic engineering for crop improvement in major crops like rice. Genetic engineering helps in the genetic manipulation of crop plants with desirable traits. For example, Ceasar et al. (2017) developed a novel *Agrobacterium*-mediated gene transformation system, and it has been successfully used to functionally characterize the *phosphate transporters 1*(*PHT1*) family genes in foxtail millet. It provides the opportunity to improve crop plants efficiently. Genetic transformation protocol is crucial for crop improvement for inserting or manipulating desirable genes. Many researchers have successfully developed an efficient *Agrobacterium*-mediated gene transformation system in finger millet (Ceasar and Ignacimuthu 2011; Satish et al. 2017). So far, many HR genes/QTLs have been identified in finger millet (Kumar et al. 2015; Singh et al. 2014; Tiwari et al. 2020). The genetic manipulation of finger millet transgenic approaches may help develop a nutrient-rich improved variety. It has some problems associated with accepting transgenic crops from regulatory bodies. Crop improvement through transgenic approaches is not safe for human health.

CRISPR/Cas system discussed above offers transgene-free engineering in crop plants. It helps to alter (remove, insert, or mutate) the desirable gene in the specific genomic regions of the targeted crop (Kim et al. 2021). Many reports have been available for improving the nutritional content in crop plants through CRISPR/Cas9 system. For example, the knock-out of the *GmFAD2–1* gene using the CRISPR/Cas9 system enhanced the oleic acid content by 16–55% in soybean seeds (Chen et al. 2011). The CRISPR/Cas9 system has successfully been used to develop transgene-free soybean plants (Haun et al. 2014; Wang et al. 2019). Application of CRISPR/ system could help identify and improve the nutrition-related genes in finger millet.

9 Brief Account of the Role of Bioinformatics as a Tool

The bioinformatic tool plays a significant role in crop improvement programs. It consists of two subfields, such as (1) computational tools and (2) databases, which help to understand living systems (Xiong 2009) better. It helps analyze and interpret various data types, including nucleotide sequences, amino acid sequences, protein structures, protein domains, etc. Bioinformatic tools and databases have provided considerable progress in plant biology by providing scientists with access to genomic information. So far, many genome-based public databases and tools are available in crop plants (Lai et al. 2012). For example, Phytozome (https://phytozome-next.jgi.doe.gov/) is an easily accessible large plant genome portal of the Department of Energy's Joint Genome Institute (DOE-JGI). It helps the plant research community access, visualize, and analyze plant genomes. It also allows mining the gene and conducting comparative genome studies. The complete and annotated genome sequence of finger millet is now available in Phytozome (https://phytozome-next.jgi.doe.gov/info/Ecoracana_v1_1). Many researchers have highlighted the use and application of bioinformatics (tools and databases) in crop improvement (Arora et al. 2018; Kushwaha et al. 2017). Recently, Ceasar's group has mined and analyzed major nutrient transporters of finger millet using various computational tools (Maharajan et al. 2022a). They collected all finger millet gene and protein sequences and analyzed the phylogeny relationship. Homology modeling was also conducted with appropriate protein templates to identify the key functional residues (Maharajan et al. 2022a). We hope the annotated finger millet genome sequence could help in the future for comparative genome analysis, gene mining, protein analysis, etc. It could be helpful for the identification of HR traits for finger millet improvement.

10 Future Prospects

Indian diet is predominantly vegetarian, based on cereals and legumes. Indian people are all accustomed to the high amount of carbohydrates and protein. Therefore, many pregnant women and older people are severely affected by Ca deficiency in India. Ca supplementation with tablets is an alternative strategy used to prevent Ca deficiency.

Although Ca tablet is cheap, straightforward, and more accessible, it suffers from side effects due to excessive Ca accumulation in vascular and soft tissues like arteries and kidneys, which may lead to heart attack and kidney stones. Finger millet grains contain more minerals such as Ca, phosphorus (P), Fe, and manganese (Mn) than other major cereals. Despite possessing a rich nutrient profile, only a few candidate genes have been identified in finger millet related to nutrient transport. No candidate genes have been characterized yet for transport function. The draft genome sequence of two different finger millet genotypes (ML-365 and PR-202) was released from 2017 to 2018. The fully annotated genome sequence is also available for mining genes and proteins. The availability of genome sequence paved the way for the GWAS to find the key genes related to nutrient fortification. Reverse genetic studies using genome editing tools like CRISPR/Cas can also be applied to finger millet to dissect the key genes involved in Ca transport and grain filling in the future. Such studies will help understand and transfer such traits into millets and non-millet cereals like rice.

11 Conclusion

Finger millet is a nutrient-dense cereal majorly grown and consumed by the poor people in Asia and Africa. The mechanism of nutrient uptake from the soil and fortification in the seeds is poorly understood. Several low resolution studies using molecular-markers have been reported for grain nutrient content before the release of the genome sequence of the finger millet. The fully annotated genome sequence of finger millet is now available, which provides the opportunity to mine and characterize the HR related genes in finger millet. This will provide the opportunity for high resolution studies with genetic and genomic tools. Genome editing tools like CRISPR/Cas may also help to this end. These studies may help to understand the HR genes and transfer or engineer such genes in other millets and non-millet cereals contributing to future crop breeding to strengthen the food security.

References

Antony U, Chandra TS (1998) Antinutrient reduction and enhancement in protein, starch, and mineral availability in fermented flour of finger millet (*Eleusine coracana*). J Agric Food Chem 46:2578–2582

Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, Liu DR (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 576:149–157. https://doi.org/10.1038/s41586-019-1711-4

Arora D, Budhlakoti N, Mishra DC, Chaturvedi KK, Kumar S, Pradhan A, Chauhan D, Singh AK, Rai A (2018) Use of bioinformatics in crop improvement. Biotech Today Int J Biol Sci 8:88

Babu BK, Agrawal PK, Pandey D, Kumar A (2014) Comparative genomics and association mapping approaches for *opaque2* modifier genes in finger millet accessions using genic, genomic and candidate gene-based simple sequence repeat markers. Mol Breed 34:1261–1279

Babu BK, Senthil N, Gomez SM, Biji KR, Rajendraprasad NS, Kumar SS, Babu RC (2007) Assessment of genetic diversity among finger millet (*Eleusine coracana* (L.) Gaertn.) accessions using molecular markers. Genet Resour Crop Evol 54:399–404

Bharathi A (2011) Phenotypic and genotypic diversity of global finger millet (*Eleusine coracana* (L.) Gaertn.) composite collection

Bollinedi H, Yadav AK, Vinod KK, Gopala Krishnan S, Bhowmick PK, Nagarajan M, Neeraja CN, Ellur RK, Singh AK (2020) Genome-wide association study reveals novel marker-trait associations (MTAs) governing the localization of Fe and Zn in the rice grain. Front Genet 11:213

Boukail S, Macharia M, Miculan M, Masoni A, Calamai A, Palchetti E, Dell'Acqua M (2021) Genome wide association study of agronomic and seed traits in a world collection of proso millet (*Panicum miliaceum* L.). BMC Plant Biol 21:1–12

Cakmak I, Hoffland E (2012) Zinc for the improvement of crop production and human health. Plant Soil 361:1–2

Ceasar A (2021) Genome-editing in millets: current knowledge and future perspectives. Mol Biol Rep 49:773. https://doi.org/10.1007/s11033-021-06975-w

Ceasar SA, Baker A, Ignacimuthu S (2017) Functional characterization of the PHT1 family transporters of foxtail millet with development of a novel *agrobacterium*-mediated transformation procedure. Sci Rep 7. https://doi.org/10.1038/s41598-017-14447-0

Ceasar SA, Ignacimuthu S (2011) Agrobacterium-mediated transformation of finger millet (*Eleusine coracana* (L.) Gaertn.) using shoot apex explants. Plant Cell Rep. https://doi.org/10.1007/s00299-011-1084-0

Ceasar S, Maharajan T, Ajeesh Krishna TP, Ramakrishnan M, Victor Roch G, Satish L, Ignacimuthu S (2018) Finger millet [*Eleusine coracana* (L.) Gaertn.] improvement: current status and future interventions of whole genome sequence. Front Plant Sci 9:1054

Chandrasekara A, Shahidi F (2010) Content of insoluble bound phenolics in millets and their contribution to antioxidant capacity. J Agric Food Chem 58:6706–6714

Chandrasekara A, Shahidi F (2012) Bioaccessibility and antioxidant potential of millet grain phenolics as affected by simulated in vitro digestion and microbial fermentation. J Funct Foods 4:226–237

Chen W, Song K, Cai Y, Li W, Liu B, Liu L (2011) Genetic modification of soybean with a novel grafting technique: downregulating the FAD2-1 gene increases oleic acid content. Plant Mol Biol Report 29:866–874. https://doi.org/10.1007/s11105-011-0286-5

Council NR (1996) Lost crops of Africa: volume I: grains. National Academies Press

Cruet-Burgos C, Cox S, Ioerger BP, Perumal R, Hu Z, Herald TJ, Bean SR, Rhodes DH (2020) Advancing provitamin a biofortification in sorghum: genome-wide association studies of grain carotenoids in global germplasm. Plant Genome 13:e20013

Dida MM, Wanyera N, Harrison Dunn ML, Bennetzen JL, Devos KM (2008) Population structure and diversity in finger millet (*Eleusine coracana*) germplasm. Trop Plant Biol 1:131–141

Doudna JA, Charpentier E (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science, 346:1258096

Fakrudin B, Shashidhar HE, Kulkarni RS, Hittalmani S (2004) Genetic diversity assessment of finger millet, Eleusine coracana (Gaertn), germplasm through RAPD analysis. PGR Newsl 138: 50–54

FAO F (2017) The future of food and agriculture-trends and challenges. Annu Rep 296:1–180

Gale MD, Devos KM (1998) Plant comparative genetics after 10 years. Science 282:656–659

Griffin GE (1974) World without cancer: the story of vitamin B17, 1st edn. CA Am Media, Los Angeles

Gull A, Jan R, Nayik GA, Prasad K, Kumar P (2014) Significance of finger millet in nutrition, health and value added products: a review. Magnes 130:120

Haun W, Coffman A, Clasen BM, Demorest ZL, Lowy A, Ray E, Retterath A, Stoddard T, Juillerat A, Cedrone F (2014) Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. Plant Biotechnol J 12:934–940

Hegde PS, Chandrakasan G, Chandra TS (2002) Inhibition of collagen glycation and crosslinking in vitro by methanolic extracts of finger millet (*Eleusine coracana*) and Kodo millet (*Paspalum scrobiculatum*). J Nutr Biochem 13:517–521

Hillary VE, Ceasar SA (2022) Prime editing in plants and mammalian cells: mechanism, achievements, limitations, and future prospects. BioEssays n/a:2200032. https://doi.org/10.1002/bies.202200032

Jaiswal V, Bandyopadhyay T, Gahlaut V, Gupta S, Dhaka A, Ramchiary N, Prasad M (2019a) Genome-wide association study (GWAS) delineates genomic loci for ten nutritional elements in foxtail millet (*Setaria italica* L.). J Cereal Sci 85:48–55

Jaiswal V, Gupta S, Gahlaut V, Muthamilarasan M, Bandyopadhyay T, Ramchiary N, Prasad M (2019b) Genome-wide association study of major agronomic traits in foxtail millet (*Setaria italica* L.) using ddRAD sequencing. Sci Rep 9:1–11

Jia G, Huang X, Zhi H, Zhao Y, Zhao Q, Li W, Chai Y, Yang L, Liu K, Lu H (2013) A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet *(Setaria italica*). Nat Genet 45:957–961

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821

Kawabata K, Yamamoto T, Hara A, Shimizu M, Yamada Y, Matsunaga K, Tanaka T, Mori H (2000) Modifying effects of ferulic acid on azoxymethane-induced colon carcinogenesis in F344 rats. Cancer Lett 157:15–21

Keiji N, Takayuki A, Nozomu Y, Satomi B, Mika K, Mayura T, Masao M, Aya M, Michihiro A, HK Y, Zenpei S, Akihiko K (2016) Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science 353:aaf8729. https://doi.org/10.1126/science.aaf8729

Kim Y-C, Kang Y, Yang E-Y, Cho M-C, Schafleitner R, Lee JH, Jang S (2021) Applications and major achievements of genome editing in vegetable crops: a review. Front Plant Sci 12:688980. https://doi.org/10.3389/fpls.2021.688980

Krishna TPA, Maharajan T, Ceasar SA (2022a) The role of membrane transporters in the biofortification of zinc and iron in plants. Biol Trace Elem Res 201:464. https://doi.org/10.1007/s12011-022-03159-w

Krishna TP, Maharajan T, Ceasar SA (2022b) Improvement of millets in the post-genomic era. Physiol Mol Biol Plants 28(3):669–685

Krishna TP, Maharajan T, Ignacimuthu S, Antony Ceasar S (2021) Genomic-assisted breeding in finger millet (*Eleusine Coracana* (L.) Gaertn.) for abiotic stress tolerance. In: Kole C (ed) Genomic designing for abiotic stress resistant cereal crops. Springer, pp 291–317, Springer Nature, Switzerland AG 2021

Krishna T, Theivanayagam M, Roch GV, Duraipandiyan V, Ignacimuthu S (2020) Microsatellite marker: importance and implications of cross-genome analysis for finger millet (*Eleusine coracana* (L.) Gaertn). Curr Biotechnol 9:160–170

Kumar A, Gaur VS, Goel A, Gupta AK (2015) De novo assembly and characterization of developing spikes transcriptome of finger millet (*Eleusine coracana*): a minor crop having nutraceutical properties. Plant Mol Biol Report 33:905–922

Kumar A, Sharma D, Tiwari A, Jaiswal JP, Singh NK, Sood S (2016) Genotyping-by-sequencing analysis for determining population structure of finger millet germplasm of diverse origins. Plant Genome 9. plantgenome2015-07

Kushwaha UKS, Deo I, Jaiswal JP, Prasad B (2017) Role of bioinformatics in crop improvement. Glob J Sci Front Res D Agric Vet 17:13–24

Lai K, Lorenc MT, Edwards D (2012) Genomic databases for crop improvement. Agronomy 2:62–73

Lee SH, Chung I-M, Cha Y-S, Park Y (2010) Millet consumption decreased serum concentration of triglyceride and C-reactive protein but not oxidative status in hyperlipidemic rats. Nutr Res 30:290–296

Maharajan T, Antony Ceasar S, Ajeesh Krishna TP, Ignacimuthu S (2021) Finger millet [*Eleusine coracana* (L.) Gaertn]: an orphan crop with a potential to alleviate the calcium deficiency in the semi-arid tropics of Asia and Africa. Front Sustain Food Syst 5:258

Maharajan T, Ceasar SA, Ajeesh krishna TP (2022b) Finger millet (*Eleusine coracana* (L.) Gaertn): nutritional importance and nutrient transporters. CRC Crit Rev Plant Sci 41:1. https://doi.org/10.1080/07352689.2022.2037834

Maharajan T, Krishna TPA, Ignacimuthu S, Ceasar SA (2022a) Finger millet genome analysis and nutrient transport. In: Kumar A, Sood S, Babu K, Gupta SM, Dayakar Rao B (eds) . Springer, The Finger Millet Genome, pp 181–199

McDonough CM, Rooney LW, Saldivar S (2000) The millets. In: Kulp K, Ponte JG Jr (eds) Handbook of cereal science and technology. Marcel Dekker Inc., New York, NY, pp 177–195

McKenzie FC, Williams J (2015) Sustainable food production: constraints, challenges and choices by 2050. Food Secur 7:221–233

Millet Network of India (2010) Millets: future of food and farming. Available at http://www.swaraj.org/shikshantar/millets.pdf

Muhu-Din Ahmed HG, Sajjad M, Zeng Y, Iqbal M, Habibullah Khan S, Ullah A, Nadeem Akhtar M (2020) Genome-wide association mapping through 90K SNP array for quality and yield attributes in bread wheat against water-deficit conditions. Agriculture 10:392

Ndlovu N, Spillane C, McKeown PC, Cairns JE, Das B, Gowda M (2021) Unravelling the genetic architecture of grain quality traits under optimum and low-nitrogen stress conditions using tropical maize (*Zea mays* L.) germplasm. Research Square. https://doi.org/10.21203/rs.3.rs-1021932/v1

Nida H, Girma G, Mekonen M, Tirfessa A, Seyoum A, Bejiga T, Birhanu C, Dessalegn K, Senbetay T, Ayana G (2021) Genome-wide association analysis reveals seed protein loci as determinants of variations in grain mold resistance in sorghum. Theor Appl Genet 134:1167–1184

Nirgude M, Babu BK, Shambhavi Y, Singh UM, Upadhyaya HD, Kumar A (2014) Development and molecular characterization of genic molecular markers for grain protein and calcium content in finger millet (*Eleusine coracana* (L.) Gaertn.). Mol Biol Rep 41:1189–1200

Panwar P, Nath M, Yadav VK, Kumar A (2010) Comparative evaluation of genetic diversity using RAPD, SSR and cytochrome P450 gene based markers with respect to calcium content in finger millet (*Eleusine coracana* L. Gaertn.). J Genet 89:121–133

Patra BC, Anilkumar C, Chakraborti M (2020) Rice breeding in India: a journey from phenotype based pure-line selection to genomics assisted breeding. Agric Res J 57:816–825

Pittas AG, Dawson-Hughes B, Li T, Van Dam RM, Willett WC, Manson JE, Hu FB (2006) Vitamin D and calcium intake in relation to type 2 diabetes in women. Diabetes Care 29:650–656

Poonia K, Chavan S, Daniel M (2012) Fixed oil composition, polyphenols and phospholipids of finger millet [*Eleusine Coracana* (L.) Gaertn]. In: Biological forum. Citeseer, pp 45–47

Pujar M, Gangaprasad S, Govindaraj M, Gangurde SS, Kanatti A, Kudapa H (2020) Genome-wide association study uncovers genomic regions associated with grain iron, zinc and protein content in pearl millet. Sci Rep 10:1–15

Puranik S, Sahu PP, Beynon S, Srivastava RK, Sehgal D, Ojulong H, Yadav R (2020) Genome-wide association mapping and comparative genomics identifies genomic regions governing grain nutritional traits in finger millet (*Eleusine coracana* L. Gaertn.). Plants People Planet 2:649–662

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152:1173–1183

Rathan ND, Krishna H, Ellur RK, Sehgal D, Govindan V, Ahlawat AK, Krishnappa G, Jaiswal JP, Singh JB, Sv S (2022) Genome-wide association study identifies loci and candidate genes for grain micronutrients and quality traits in wheat (*Triticum aestivum* L.). Sci Rep 12:1–15

Ray S, Satya P (2014) Next generation sequencing technologies for next generation plant breeding. Front Plant Sci 5:367

Ruanjaichon V, Khammona K, Thunnom B, Suriharn K, Kerdsri C, Aesomnuk W, Yongsuwan A, Chaomueang N, Thammapichai P, Arikit S (2021) Identification of gene associated with

sweetness in corn (*Zea mays* L.) by genome-wide association study (GWAS) and development of a functional SNP marker for predicting sweet corn. Plants 10:1239

Salimath SS, de Oliveira AC, Bennetzen JL, Godwin ID (1995) Assessment of genome origins and genetic diversity in the genus Eleusine with DNA markers. Genome 38:757–763

Sands DC, Morris CE, Dratz EA, Pilgeram AL (2009) Elevating optimal human nutrition to a central goal of plant breeding and production of plant-based foods. Plant Sci 177:377–389

Satish L, Ceasar SA, Ramesh M (2017) Improved *agrobacterium*-mediated transformation and direct plant regeneration in four cultivars of finger millet (*Eleusine coracana* (L.) Gaertn.). Plant Cell Tissue Organ Cult 131:547–565. https://doi.org/10.1007/s11240-017-1305-5

Sharma D, Jamra G, Singh UM, Sood S, Kumar A (2017) Calcium biofortification: three pronged molecular approaches for dissecting complex trait of calcium nutrition in finger millet (*Eleusine coracana*) for devising strategies of enrichment of food crops. Front Plant Sci 7:2028

Sharma D, Tiwari A, Sood S, Jamra G, Singh NK, Meher PK, Kumar A (2018) Genome-wide association mapping of agro-morphological traits among a diverse collection of finger millet (*Eleusine coracana* L.) genotypes using SNP markers. PLoS One 13:e0199444

Sharma D, Tiwari A, Sood S, Meher PK, Kumar A (2022) Identification and validation of candidate genes for high calcium content in finger millet [*Eleusine coracana* (L.) Gaertn.] through genome-wide association study. J Cereal Sci 107:103517. https://doi.org/10.1016/j.jcs.2022.103517

Shi X, Zhou Z, Li W, Qin M, Yang P, Hou J, Huang F, Lei Z, Wu Z, Wang J (2022) Genome-wide association study reveals the genetic architecture for calcium accumulation in grains of hexaploid wheat (*Triticum aestivum* L.). BMC Plant Biol 22:1–14

Shobana S, Krishnaswamy K, Sudha V, Malleshi NG, Anjana RM, Palaniappan L, Mohan V (2013) Finger millet (Ragi, *Eleusine coracana* L.): a review of its nutritional properties, processing, and plausible health benefits. Adv Food Nutr Res 69:1–39. https://doi.org/10.1016/B978-0-12-410540-9.00001-6

Singh UM, Chandra M, Shankhdhar SC, Kumar A (2014) Transcriptome wide identification and validation of calcium sensor gene family in the developing spikes of finger millet genotypes for elucidating its role in grain calcium accumulation. PLoS One 9:e103963

Song G, Jia M, Chen K, Kong X, Khattak B, Xie C, Li A, Mao L (2016) CRISPR/Cas9: a powerful tool for crop genome editing. Crop J 4:75–82

Sood S, Kumar A, Babu BK, Gaur VS, Pandey D, Kant L, Pattnayak A (2016) Gene discovery and advances in finger millet [*Eleusine coracana* (L.) Gaertn.] genomics – an important Nutri-cereal of future. Front. Plant Sci 7:1634

Subba Rao M, Muralikrishna G (2002) Evaluation of the antioxidant properties of free and bound phenolic acids from native and malted finger millet (Ragi, *Eleusine coracana* Indaf-15). J Agric Food Chem 50:889–892

Swaminaidu N, Ghosh S, Mallikarjuna K (2015) Millets: the miracle grain. Int J Pharm Bio Sci 6: 440–446

Tahir ul Qamar M, Faryad A, Bari A, Zahid B, Zhu X, Chen L-L (2020) Effectiveness of conventional crop improvement strategies vs. omics. In: Fahad S, Hasanuzzaman M, Alam M, Ullah H, Saeed M, Khan IA, Adnan M (eds) Environment. Climate, Plant and Vegetation Growth. Springer, pp 253–284. Springer Nature Switzerland AG 2021

Thapliyal V, Singh K (2015) Finger millet: potential millet for food security and power house of nutrients. Int or Res Agric For 2:22–23

Tiwari A, Sharma D, Sood S, Jaiswal JP, Pachauri SP, Ramteke PW, Kumar A (2020) Genome-wide association mapping for seed protein content in finger millet (Eleusine coracana) global collection through genotyping by sequencing. J Cereal Sci 91:102888

Upadhyay S, Kumar AR, Raghuvanshi RS, Singh BB (2011) Nutritional status and knowledge of hill women on anemia: effect of various socio-demographic factors. J Hum Ecol 33:29–34

Upadhyaya HD, Gowda CLL, Pundir RPS, Reddy VG, Singh S (2006) Development of core subset of finger millet germplasm using geographical origin and data on 14 quantitative traits. Genet Resour Crop Evol 53:679–685

Van K, Rastogi K, Kim KH, Lee SH (2013) Next-generation sequencing technology for crop improvement. SABRAO J Breed Genet 45:84–99

Verma RK, Chetia SK, Sharma V, Baishya S, Sharma H, Modi MK (2022) GWAS to spot candidate genes associated with grain quality traits in diverse rice accessions of North East India. Mol Biol Rep 49:5365–5377

Vetriventhan M, Azevedo VCR, Upadhyaya HD et al (2020) Genetic and genomic resources, and breeding for accelerating improvement of small millets: current status and future interventions. Nucleus 1:217–239. https://doi.org/10.1007/s13237-020-00322-3

Wang S, Yokosho K, Guo R, Whelan J, Ruan Y-L, Ma JF, Shou H (2019) The soybean sugar transporter GmSWEET15 mediates sucrose export from endosperm to early embryo. Plant Physiol 180:2133–2141

Welch R (2002) The impact of mineral nutrients in food crops on global human health. Plant Soil 247:83–90

WHO (2016) Guideline daily iron supplementation in infants and children. World Health Organization

Xiong J (2009) Essential bioinformatics: introduction and biological databases. Cambridge University Press. New York, USA

Yadav S, Kumar A, Sood S (2020) Unraveling the genetics of calcium content in finger millet grains through association mapping. Indiam J Genet 80:432–440

Yadav CB, Srivastava RK, Gangashetty PI, Yadav R, Mur LAJ, Yadav RS (2021) Metabolite diversity and metabolic genome-wide marker association studies (Mgwas) for health benefiting nutritional traits in pearl millet grains. Cell 10:3076

Zhang N, Wang M, Fu J, Shen Y, Ding Y, Wu D, Shu X, Song W (2020) Identifying genes for resistant starch, slowly digestible starch, and rapidly digestible starch in rice using genome-wide association studies. Genes Genom 42:1227–1238

Zhao M, Lin Y, Chen H (2020) Improving nutritional quality of rice for human health. Theor Appl Genet 133:1397–1413

Zheng Y, Yuan F, Huang Y, Zhao Y, Jia X, Zhu L, Guo J (2021) Genome-wide association studies of grain quality traits in maize. Sci Rep 11:1–12

Proso Millet Nutraceutomics for Human Health and Nutritional Security

Rituraj Khound, Ramesh Kanna Mathivanan, and Dipak K. Santra

Contents

R. Khound · R. K. Mathivanan · D. K. Santra (✉)
Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Scottsbluff, NE, USA
e-mail: rkhound@huskers.unl.edu; rmathivanan2@huskers.unl.edu; dsantra2@unl.edu

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_10

Abstract

Proso millet (*Panicum miliaceum* L.) is a climate-resilient, ancient, and gluten-free cereal. This allotetraploid crop is healthy for humans and the environment. Its use as human food is steadily increasing, especially for people with diabetes and celiac disease due to its exceptional nutritional properties and gluten-free starch. There is no comprehensive review on proso millet seed nutraceuticals quantity, quality, characteristics, genetics, genomics (omics), and genetic improvement strategies. This review article aims to summarize published research on these aspects of proso millet seed nutrients and nutraceuticals. There are lots of reports on proso millet seed nutrients (macro, micro, and secondary metabolites) and their health benefits. There are a significant number of resources of "omics" tools for proso millet overall genetic improvement. However, no or very little genetic and genomic information (e.g., genes and genetic control mechanisms) and biosynthesis of majority of the nutraceuticals of proso millet seed are available. It may take years to see the full potential of "omics" for proso millet nutraceutomics. Phytochemical analyses and omics technologies are getting more efficient, faster, and cheaper. With this new opportunity, it is essential that proso millet scientists around the world especially food technologists, plant breeders, and geneticists of both public and private sectors work collaboratively for the genetic improvement of proso millet for nutraceutical. This will stimulate industries to use more proso millet in food products for human health and nutritional security under global climate change.

Keywords

Climate-resilient · Gluten-free · Health-promoting · Omics · GWAS · Genomic selection · Nutraceuticals

1 Introduction

Proso millet (*Panicum miliaceum* L.), a small-grain, annual cereal, belongs to the *Panicoideae* subfamily of the Poaceae family. It is one of the earliest cereal crops domesticated by mankind. It was reported that proso millet was domesticated around 8000 BC in the semi-arid regions of northern China (Lu et al. 2009). This minor millet is known by different names in different geographical locations. The word "proso" is a pan-Slavik name for millet (Santra et al. 2019). In the USA, it is

commonly referred to as common millet or hog millet. Proso millet is also known as broomcorn millet in China; common millet in Japan, Korea, and other Asia-Pacific countries; Gijang in Korea; Barri in India; Mijo in Spain; Hersey millet in Germany; and French white or panic millet in France (Santra et al. 2019; Das et al. 2019). This millet is globally distributed owing to its wide adaptability to different climatic zones. The most notable countries producing proso millet are China, the USA, India, Ukraine, Russia, Nepal, Belarus, Turkey, and Romania (Das et al. 2019) (Fig. 1). Proso millet seeds were brought to the USA in 1875 by German-Russian immigrants who first cultivated this cereal along the eastern Atlantic coast of the country. Its cultivation later spread westward into the interior of the North American continent (Wietgrefe 1990; Santra 2013; Habiyaremye et al. 2017).

Proso millet plants are erect, 30–100 cm tall with hollow stems (known as *culm*) and an adventitious root system (Baltensperger 2002). Proso millet plants develop 10–45 cm long, drooping panicles which can be either open (*lax*) or compact. The grains (botanically known as *caryopsis*) are normally 3 mm long and 2 mm wide. A great deal of variation is observed in the panicle's shape and grain color. The panicle morphology can vary widely from its shape (e.g., compact, open, semi-open), attitude (e.g., erect, semi-erect, dropping), and branches (e.g., lax, medium, and dense). The grains can be white, cream, yellow, red, orange, brown, green, black, or tan (Santra et al. 2019). Proso millet is predominantly self-pollinated with ~10% cross-pollination. It is considered an allotetraploid with nine basic sets of chromosomes ($2n = 4x = 36$). The proso millet genome is estimated to be 930 megabase pairs (Mbp) (Zou et al. 2019). Hunt et al. (2014) studied the evolution of proso millet species. Proso millet was believed to have evolved through the natural hybridization of two diploid progenitors followed by the natural polyploidization. Hunt et al. reported that diploid witchgrass (*Panicum capillare* L.) and one diploid parent of the

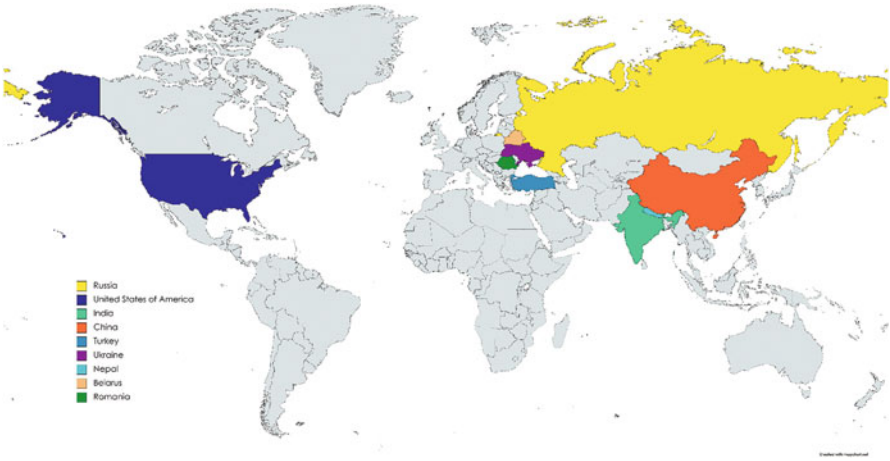


Fig. 1 Major proso millet growing countries in the world. (The map was created on mapchart.net (https://www.mapchart.net/world.html))

tetraploid torpedo grass (*P. repens*) are the probable diploid parents of proso millet (Hunt et al. 2014).

Proso millet is primarily cultivated in the predominantly winter wheat-growing semi-arid areas of the High Plains of the USA (Rajput et al. 2016). This region receives an average annual precipitation of ~400 mm. Conserving soil by minimizing water loss and maintaining high organic matter is paramount to sustaining agriculture in this dry region. The usual practice of summer fallow to preserve soil moisture often leads to depletion of soil organic matter (Nielsen and Calderón 2011). Planting wheat following corn and millet and only millet reportedly doubled the productivity in comparison with the wheat-fallow system (Anderson et al. 1999). Even though a wide selection of crops, such as corn, teff, millets, soybean, and peas, are available for crop rotation, proso millet stands out as a rotational crop for the dryland production system. This is because proso millet, being a C4 crop, has exceptional water-use efficiency and can survive harsh conditions like drought, high temperatures, and limited available nutrients (Habiyaremye et al. 2017). This cereal has a short-growing season (60–90) days, which allows it to avoid prolonged exposure to drought conditions. Besides, the shallow root system (90–120 cm) conserves deep soil moisture for the next crop (Das et al. 2019). Cultivation of proso millet lowers weed, disease, and insect pressure on wheat (Habiyaremye et al. 2017). All these attributes make proso millet an excellent choice as a rotation crop in rainfed agricultural systems under water-stress situation.

Millet is considered a nutraceutical and functional food because of its superior seed nutrients such as starch, proteins, vitamins, minerals, dietary fibers, and antioxidant activities. It also has other bioactive compounds such as resistant starch, lipids, phenolic acid, flavonoids, lignin, phytosterols, phytic acid, and tannins present in millets that have health-promoting properties (Tripathi et al. 2021). Proso millet is gluten-free and is abundant in vitamin B, iron, calcium, potassium, zinc, and magnesium (Das et al. 2019). Proso millet could be exploited as nutraceuticals and used in therapeutic applications for chronic diseases such as obesity, cardiovascular diseases, cancer, and diabetes (Majid and Priyadarshini 2020). Millets are grossly neglected as human food despite being highly nutritious primarily due to a lack of awareness. The increasing number of reports on health benefits of millets in recent years generated a renewed interest in using millets for biomedical research. Proso millet grain has protective and preventive effects against old age-related diseases such as diabetes, high blood pressure, cardiovascular diseases, cancer, Parkinson's disease, and other metabolic syndromes (Kalinová 2007).

The use of proso millet grain for human consumption predates recorded human history. This small grain was widely used for preparing traditional sweet and savory cuisines in Russia, China, and Germany thousands of years ago (Santra and Rose 2013). Proso and other millet-based food are quite common in many countries in Asia and Africa. However, this ancient grain is largely under-utilized in developed countries in Europe and North America despite its remarkable nutritional and health-promoting properties (Das et al. 2019). Proso millet grains are primarily used as birdseed and animal feed in the USA and Europe (Santra et al. 2019). Nevertheless, these grains are primarily used for human consumption in several countries

including China, Korea, Japan, India, and Russia. Many traditional cuisines and commercial food products are prepared from whole millet grains. Proso millet grains are used in porridge, rice-based meals, noodles, bread, pasta, flour, pancakes, and alcoholic beverages (Santra et al. 2019; Das et al. 2019). Proso millet grains can be used for ethanol production like corn (Rose and Santra 2013).

The term "omics" is commonly used to refer to a broad field of science and engineering focusing on the interactions among biological information objects pertaining to various "omes" (https://omicstutorials.com/home/). It involves collective quantification and characterization of collections of biological molecules that define the structural and functional attributes of an organism. The primary focuses of omics are on (1) mapping information objects such as genes, proteins, and ligands, (2) finding interaction relationships among the objects, (3) engineering the networks and objects to understand and manipulate the regulatory mechanisms, and (4) integrating various "omes" and "omics" subfields. The five major types of "omics" are (1) genomics, which deals with the analysis of the DNA sequence; (2) transcriptomics, which deals with the analysis of transcribed RNA; (3) proteomics, which focuses on proteins present in a sample; (4) metabolomics, which involves identification and quantification of all metabolites in a sample; and (5) miRNAomics, which studies regulatory mechanisms underlying control of transcription of genes. Here, we introduced a new term called "nutraceutomics" to describe the comprehensive study associated with all phytochemicals present in seeds of food crops (proso millet in this chapter) and their effects on human health and nutrition.

The goal of this chapter is to provide a comprehensive review of available literature for (1) proso millet seed composition and their impacts on human health, (2) variabilities of the seed compositions within proso millet germplasm available across the globe, and (3) genetic improvement for changing proso millet seed compositions through breeding (conventional and modern) and biotechnologies.

2 Proso Millet Seed Nutraceuticals

A nutraceutical is defined as any substance that is a food or part of a food and provides health benefits, including the prevention and treatment of disease. The term nutraceutical is the hybrid of "nutrition" and "pharmaceutical" and was coined by Stephen Defelice in 1989 (Kalra 2003). Nutraceuticals can be of two types: (1) macronutrients such as carbohydrates, proteins, lipids, and fibers (Table 1) and microelements such as minerals (Table 2) and vitamins (Table 3) and (2) secondary metabolites such as polyphenols, flavonoids, etc. These nutraceuticals in proso millet are discussed below.

2.1 Starch

Starch, a polysaccharide, is the major portion (97.1%) of the total saccharides (64.5 g/100 g) in proso millet (Kalinová 2007). Like in any other cereal, there are

Table 1 Macro nutrients composition per 100 g of proso millet and other millets

Source	Carbohydrates (g)	Crude Protein (g)	Fat (g)	Crude fiber (g)	Ash (g)	Energy (cal)	% Of daily calories (2200 for adult men and 1800 for adult women by Ladabaum et al. 2014)	
							Adult men	Adult women
Proso millet	**55.0–70.0**	**10.0–13.0**	**1–3.5**	**2.0–9.0**	**2.0–4.0**	**330–340**	**15–16**	**18–19**
Little millet	60.0–75.0	10.0–15.0	5.0–6.0	4.0–8.0	2.5–5.0	329–341	15–16	18–19
Pearl millet	60.0–76.0	12.0–14.0	4.8–5	7.2–2.5	2.0–2.2	363–412	17–18	20–21
Finger millet	60.0–80.0	7.0–10.0	1.3–1.8	3.6–4.2	2.6–3.0	328–336	15	18–19
Foxtail millet	59.0–70.0	11.2–15.0	4.0–7.0	4.5–7.0	2.0–3.5	330–350	15–16	18–20
Kodo millet	66.0–72.0	8.0–10.0	1.4–3.6	5.0–9.0	4.0–5.0	309–353	14–16	17–20
Barnyard millet	55.0–65.0	6.0–13.0	2.0–4.0	9.5–14.0	4.0–4.5	300–310	14–15	16–17

Note: The data was adapted from Tripathi et al. (2021)

Table 2 Mineral composition per 100 g of proso millet and other millets

Minerals (mg/100 g)	K	Na	Mg	Ca	P	Mn	Zn	Cu	Fe
Proso millet	**250–320**	**8.2–10**	**117–153**	**20–23**	**230–281**	**0.6–1.81**	**1.4–2.4**	**0.83–5.8**	**4.0–5.2**
Little millet	129–370	6–8.1	120–133	12.0–31.0	215–310	1–20	3.5–11	1–4	13–20
Pearl millet	440–442	10.0–12.0	130–137	10.0–46.0	350–379	1.15–1.8	2.95–3.1	0.62–1.06	7.49–8.0
Finger millet	408–570	7.0–11.0	110–137	240–410	240–320	5–5.5	2–2.3	0.4–4	3.9–7.5
Foxtail millet	250–400	4.6–10	100-130	10.0–30.0	270–310	2.19–26	2.14–9	1–3.0	3.26–19
Kodo millet	144–170	4.6–10	130–166	10.0–31.0	215–310	1.10–2.9	0.7–1.5	1.6–5.8	0.7–3.6
Barnyard millet	–	–	–	–	–	–	–	–	–

Note: The data was adapted from Tripathi et al. (2021)

Table 3 Vitamin composition per 100 g of proso millet and other millets

Vitamins (mg)	Thiamine (B1)	Riboflavin (B2)	Niacin (B3)	Pyridoxin (B6)	Pantothenic acid (B5)	Biotin (B7)
Proso millet	**0.41**	**0.22**	**1.55**	**0.52**	–	–
Little millet	0.26	0.05	1.29	0.04	0.60	6.03
Pearl millet	0.25	0.20	0.86	0.27	0.50	0.64
Finger millet	–	–	–	–	–	–
Foxtail millet	0.59	0.11	3.20	–	0.82	–
Kodo millet	0.29	0.20	1.49	0.07	1.49	
Barnyard millet	0.33	0.10	4.20	–	–	–

Note: The data was adapted from Rao et al. (2017)

two types of starch in proso millet, namely, amylopectin and amylose. Amylose is the straight-chain polymer of glucose, whereas amylopectin is the branched-chain polymer of glucose units. In proso millet, the amylose content varies from 17.21% to 32.6% (dry basis), and amylopectin varies from 67.40% to 82.79%. The amylose content in proso millet starch is a little higher than maize (17–27%) and marginally lower than rice (28%). Native starch of proso millet has the highest digestibility (50%) in comparison to other millets (Kumari and Thayumanavan 1998), and the starch digestibility is similar to that of maize (43%). In general, grain starch has very low solubility in water. The proso millet starch is approximately 6.89% soluble at 90 °C. Moreover, millet starch exhibit a higher water-binding capacity than wheat starch and has a higher tendency to break down (~50 Brabender units - BU). Proso and foxtail millets show lower set back values (about 330 BU)compared to other millets (Kumari and Thayumanavan 1998).

The dietary patterns of many people concerned about healthy eating habit are changing across the world. Proso millet is gradually gaining attention owing to its impressive nutrient profile and gluten-free property. The eating quality and cooking parameters of proso millet grains are attributable to the starch properties. The starch granules in proso millet vary greatly in size and shape. The size of the granules range between 0.3 and 17 μm and can be round or polygonal in shape. Proso millet varieties are categorized as waxy (~1% amylose) and non-waxy (~25% amylose) types. Another important reason for consuming the proso millet is its high resistant starch content, which causes slow starch digestion, particularly in non-waxy varieties. Because of these reasons, the non-waxy starch varieties of proso-millet are widely used in food processing. The variation in amylose content and structural integrity contribute to the pasting and thermal behavior of proso millet starch. The starch of proso millet showed some promising functional properties that make it a suitable alternative to commercially available starch for food and non-food

applications. There are several reports on the use of proso millet seeds in the preparation of different food products like gluten-free bread, porridge, pasta, ready-to-eat breakfast cereals, infant foods, and distilleries (Das et al. 2019). It can be concluded that proso millet is an alternative source of starch considering its quality characteristics (Bangar et al. 2021).

2.2 Protein

The seed protein content of proso millet is comparable to maize and wheat. Proso millet seeds contain higher amount of protein than any other millets (Geervani and Eggum 1989). The protein content depends on the variety, environmental conditions, soil nutrients, and weather conditions during grain formation. The average seed protein in proso millet ranges from 11.5% to 13.0% with a maximum of about 17% (Kalinova and Moudry 2006). Varieties with a red seed coat have the lowest quantity and quality of protein (Essential Amino Acid Index, or EAAI) than light-colored grains. Dehulled grains have a little bit higher protein content (12.3–16.3%) due to the removal of the seed coat (hull) that is poor in proteins (Jones et al. 1970; Ravindran 1992). The protein quality (EAAI) of proso millet grain (51–67%) is higher than that in maize.

Proteins are complex biomolecules as they may sometimes comprise several subunits with various physicochemical and functional properties. On the basis of solubility, plant storage proteins are classified as albumin (water-soluble), globulin (saline-soluble), prolamin (alcohol-soluble), and glutelin (alkaline-soluble). Prolamin (31–50%) is the major storage protein in proso millet like in most cereals (Jones et al. 1970; Dendy 1995). Glutinous (waxy) proso millet contains a higher proportion of albumins and globulins (18–31%) than non-glutinous types (13–16%) (Parameswaran and Thayumanavan 1995). Similar to other cereals, proso millet grain protein has inadequate amount of content (3.68%). However, proso millet protein is rich of some essential amino acids (leucine, isoleucine, methionine) compared to wheat protein. Millet protein is believed to have beneficial influence on cholesterol metabolism. Proso millet is well-suited for patients on a gluten-free diet as its grains contain the specific prolamin fraction under the permitted level (Kalinova and Moudry 2006).

The proximate analysis of seed storage proteins is essential for understanding the basic physical and functional properties of proso millet cultivars to form the foundation for developing value-added products. Nine US proso millet varieties were characterized for proximate analysis by Singh et al. (2018). They observed a wide range of values for the physical properties, such as sphericity (0.86–0.91), volume 3.94 to 5.14 mm^3, bulk density (765.49–809.67 kg m^{-3}), porosity (42.49–44.20%), and angle of repose (22.98–25.74°). The varieties were also evaluated for their pasting and gelatinization properties. A high correlation was observed between amylose content and onset temperature ($r = -0.94$), peak gelatinization temperature ($r = -0.92$), peak viscosity ($r = 0.84$), final viscosity ($r = 0.91$), and setback viscosity ($r = 0.90$) (Singh et al. 2018).

The knowledge of the protein structure-function relationship is necessary to understand the nutritional properties of seed storage proteins. The physicochemical, functional, and structural properties of the storage protein fraction from two defatted proso millet cultivars, viz., Dawn and Plateau, were determined by Akharume et al. (2020). The protein recovery efficiency for the two varieties were determined to be 53.5% (Dawn) and 60.1% (Plateau). The average denaturation temperature of all fractions was 82.1 ± 3.5 °C. Surface hydrophobicity values of Dawn were 11,781, 10,594, 316, and 2225 for albumin, globulin, and glutelin, respectively, and 3415, 2865, 353, and 456 of Plateau fractions, respectively. Most of the protein fractions exhibited the highest solubility at pH 9 and lowest solubilities at pH $\leq$ 7. The solubility varied from 5.7% to 100%. Emulsifying activity index (EAI) of less than 0.25 m^2/g was estimated for most fractions, while the highest emulsion stability index (ESI) was about 60 min. Prolamin fraction showed three major peptide bands of 11, 14, and 24 kDa, while glutelin fraction revealed only a major band of 15 kDa and several minor bands of 11, 22, 24, 78, and 209 kDa. No differences in the electrophoresis pattern were observed for the fraction with or without a reducing agent (Akharume et al. 2020).

Digestibility of seed proteins of proso millet and other millets is reduced significantly after cooking due to conformational changes. Panicin, zein-like, the major seed storage protein fraction in proso millet, produced hydrophobic aggregates upon heating that are extremely resistant to protease digestion (Gulati et al. 2017). Using in vitro digestion as a measure of digestibility, aggregation reduces digestibility from about 80% to <40%. The recalcitrant hydrophobic protein aggregates that form as a result of heating are resistant to both aspartic proteases (pepsin) and serine proteases (chymotrypsin/trypsin). To estimate variation in "digestibility" as a trait, in vitro digestibility across 33 accessions of proso millet from different countries was measured and observed significant variation, ranging from 26% to 57% digestibility after heating (Gulati et al. 2018). Thus, there appears to be ample trait variation for selective breeding.

To further understand the nature of this digestibility trait, liquid chromatography-mass spectrometry (LC-MS)/MS analysis was done on the lines with high and low protein digestibility and identified three peptide sequences from the panicin fraction that were correlated with low digestibility and two that were associated with high digestibility (Gulati et al. 2018). Three peptide sequences of Gulati et al. (2018) were used to query annotated genes *Panicum miliaceum* genome (Zou et al. 2019). Seven different genes were identified belonging to the alpha family of zein proteins, the major seed storage proteins of maize. BLAST alignments identified a total of 14 members of the zein-alpha family in proso millet based on alignments showing >35% identity to the query. Phylogenetic analysis (Fig. 2a) further divided these zein-like genes into five subfamilies that are related to the PMS-1, P22.1/22A-1, and ZA1/M1 members of the 19 KDa and 22KDa zein alpha family and two members of the 10KDa prolamin-like protein family. The arrangement of the 14 zein-like proteins and the prolamin proteins on the proso millet genome (Fig. 2b) revealed an architecture suggesting that two pairs of highly related zein-like genes have arisen through chromosome duplications (Chr3-Chr4 and Chr9-Chr10), while a combination of five highly related 22 KDa zein-like genes on Chr17-Chr18 appears to have arisen from a

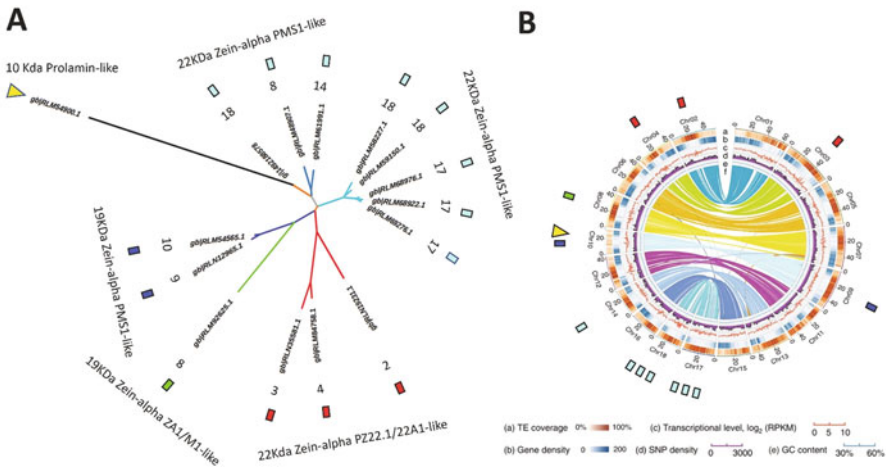


Fig. 2 Panicin (zein-like), prolamin gene family in proso millet (*Panicum miliaceum*). (**a**) Genes identified using three peptide sequences of prolamin fraction (Gulati et al. 2017) in BLAST alignment of proso millet genome sequence (Zou et al. 2019). Fourteen zein-line genes are divided into five sub-families. Chromosome number is indicated on top of each gene (unpublished). (**b**) The proso millet genome sequence assembly reported by Zou et al. (2019). This image, licensed under a Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), was used with minor modifications.

combination of ancestral gene duplications within the progenitor of the Chr17/Chr18 chromosomes, followed by duplication of the ancestral chromosome. The four remaining genes across Chr2, Chr8, Chr14, and Chr18 form distinct families that are more distantly related. Structurally, the amino acid composition of predicted proteins from each of these zein-like subfamilies shares features common to zein proteins, including enrichment in polar amino acids, and 9–10 different stretches of repeated sequence containing poly-glutamine (QQQ) separated on average by 11 residues. Structural analysis of the maize zeins has shown the repeated sequences form alpha helices with the polar amino acids (including glutamine residues) interacting with residues on adjacent anti-parallel helices to form a bundle of multiple alpha helices in a compact structure (Argos et al. 1982; Matsushima et al. 1997).

2.3 Lipids

Dehulled grains of proso millet contain 3.5% to 6.7% of lipid (Jones et al. 1970; Ravindran 1992; Kalinová 2007). Among cereals, only oats have a higher (7.14%) lipid than that in proso millet. The germ contains about 25% of the total lipids. Proso millet gains contain 62.2% of free, 27.8% of bound, and 10% of structural lipids. Unsaturated fatty acids make up 85–89% of proso millet lipids of which 42% are polyunsaturated acids (Kalinová 2007). Polyunsaturated lipids (particularly linoleic acid) play an essential role in cholesterol metabolism. The major fatty acids of proso

millet are linolenic (38.4–66.68%), oleic (21.4–22.7%), and palmitic (6.61–11.3%) acids. The physicochemical characteristics of proso millet lipids (consistence, setting point, acid number) are similar to the lipid content of sunflower, maize, and soya (Kalinová 2007).

2.4 Fiber

Cereal grains serve as the most significant source of dietary fiber for the human diet. The fiber content in the human diet is critical for the prevention and treatment of high blood pressure and high cholesterol. The fiber content in dehulled grains of proso millet is like that of oat (0.8–1.2%). The composition of cell wall in proso millet grains is similar to that of tissues of other grass species. The cell wall comprises xyloglucan, arabinoxylan, uronic acid, arabi-nosyl, galactosyl residues, arabinogalactan, and β-D-glucans (Kalinová 2007). The β-D-glucans (glucose units linked to a long polymer chain) play an important role in human nutrition by decreasing total blood cholesterol. The β-glucans content in proso millet (0.5–1.0%) is at the same level of lentils (0.4–1.1%), maize (0.5–1.3%), rice (0.4–0.9%), spelt (0.6–1.2%), or wheat (0.5–1.0%) (Demirbas 2005).

2.5 Minerals

The majority of proso millet seed mineral is present in its pericarp, aleurone layer, and germ. The mineral content varies from 1.5% to 4.2% (Ravindran 1992; Kalinová 2007), which is higher than that in wheat grain (1.5–2.0%). The majority of minerals in proso millet are slightly higher or similar to those of other cereals (Table 2). The common food processing and dehulling reduces the mineral content to around 27–53% (Kalinová 2007). This reduction commensurate with the dehulling intensity. Proso millet is deficient in calcium, but the grain contains a high amount of phosphorus. However, the bondage of phosphorus with phytates considerably interferes with its bioavailability. Proso millet is abundant in potassium, iron, and manganese (19.5–20.6 mg/100 g) (Ravindran 1992). Its grain is almost devoid (0.01 g/100 g) of sodium. Proso millet is an excellent source of zinc, copper, and boron as well (11.6 mg/kg) (Kalinová 2007).

2.6 Vitamins

The dehulled grain of proso millet is rich in vitamin B1, thiamine (0.42–0.80 mg/100 g); B2, riboflavin (0.22–0.40 mg/100 g); B3, niacin (1.55–3.7 mg/100 g); B6, pyridoxin (0.52–0.80 mg/100 g); and E, tocopherol (0.1–2.60 mg/100 g) (Table 3). The levels of vitamins B1 and B2 are twice that of rice, wheat, or barley. Most of the vitamins are deposited in the germ and aleurone layer and therefore, decortication considerably decreases their content. The color of dehulled grains is correlated to carotene and xanthophyll content. The carotene content is relatively stable and does not vary due to

growing conditions such as soil type, level of fertilization, and interannual weather changes (Kalinová 2007). The unrefined fat extract of proso millet grain contains 8.3 to 10.5 mg of vitamin A and 87 to 96 mg of vitamin E per 100 g (Tripathi et al. 2021).

2.7 Antinutritional Factors

Knowledge of antinutritional factors such as tannins, phytates, oxalates, and enzyme inhibitors is essential to predict the nutritional value of products. Phytates affect the bioavailability of certain minerals, particularly multivalent cations including P, Ca, Mg Fe, and Zn, by binding with them. Phytic acid forms a complex with most of the free P (~67.3%) in the millets (Ravindran 1992). The phytic acid content in proso millet varies from 0.17 to 0.61 g/100 g, which is higher than that in polished rice but lower than that in wheat. Dehulling reduces the phytate content by 17–24% and thus makes more P bioavailable (Lorenz 1983).

3 Secondary Metabolites in Proso Millet

Like other millets and cereals, proso millet grain contains many secondary metabolites, which reportedly have beneficial effects on human health. The following section summarizes the secondary metabolites in proso millet.

3.1 Phenolic Compounds

Polyphenols are the major group of phytochemicals in plants, and they have numerous health benefits. Polyphenols in cereals have an adverse influence on color, flavor, and nutritional quality. These compounds are mostly localized in the aleurone layers of the grain. The polyphenol content in plants is influenced by factors such as plant part, developmental stage, and environment. Total phenolic compounds in proso millet are about 0.05–0.10 mg per 100 g of catechin equivalents, on dry basis. The bound polyphenols (1% HCl extractable) in proso millet were (2.21 ± 0.01), which was significantly lower than that in other millets such as kodo millet (81.64 ± 0.15), foxtail millet (11.59 ± 0.23), little millet (9.64 ± 0.28), pearl millet (9.14 ± 0.17), and finger millet (3.83 ± 0.18). Bound phenolic compounds have antioxidant, anti-obesity, anti-diabetic, antimutagenic, anticarcinogenic, antimicrobial, and antiviral properties (Tripathi et al. 2021).

3.2 Flavonoids

Flavonoids are a group of plant secondary metabolites with a general structure of a 15-carbon skeleton. Flavonoids such as catechin, quercetin, anthocyanin, tannin, etc. have human health-promoting properties, owing to their pharmacological activities

such as radical scavenging. There are no reports of flavonoids in proso millet. However, a few reports are there for other millets. Finger millet leaves are known to contain eight types of flavones: vitexin, isovitexin, saponarin, violanthin, orientin, iso-orientin, lucenin-1, and tricin. Pearl millet contains glucosylvitexin, glucosylorientin, and vitexin in a ratio of 29:11:4 (Kalinová 2007). Finger millet is reportedly the only millet that contains condensed tannins.

3.3 Carotenoids

Carotenoids are well-known for their provitamin A activity. Carotenoids prevent various diseases because they act as antioxidants. A recent report showed that the carotenoid content in edible millet flour varied from 78 to 366 μg/100 g with an average of 366, 199, 78, 173, and 366 μg/100 g in proso, finger, little, and foxtail millets, respectively (Asharani et al. 2010). The carotenoid contents of millets were comparable to that of wheat (150–200 μg/100 g) and sorghum (180–230 μg/100 g) but significantly lower than that in maize (1800–5500 μg/100 g) (Tripathi et al. 2021).

3.4 Phytosterols

Phytosterols, which are desmethyl sterols, are essential structural and functional components of plant cells. Phytosterol esters can potentially reduce the blood serum LDL (low-density lipoprotein) cholesterol levels down to 14% but do not affect HDL (high-density lipoprotein) levels. The risk of heart disease can be reduced by up to 40% by including proso millet in the regular diet depending on age and some other factors. Other than finger millet, there are no reports of phytosterols in proso and other millets. The sterol content of finger millet was reported to be 0.149% on a seed weight basis, whereas other millets contain only a trace amount. Phytosterol content of sorghum and corn was estimated to be 0.5 mg/g and 0.9 mg/g (Singh et al. 2003).

4 Proso Millet Grain for Human Health

Proso millet confer many beneficial effects to human health. Some of these effects are backed by more scientific evidence than others. Several epidemiological studies have shown that regular consumption of millets and their products can ameliorate the affect of chronic diseases such as diabetes, cardiovascular disease, cancers, and all-cause mortality (Table 4).

4.1 Antioxidant

Millets including proso millet are abundant in both soluble and insoluble bound polyphenols, which have high antioxidant potential with metal chelating and

Table 4 Human health benefits and contributing nutraceuticals of proso millet

Human health benefits	Nutraceutical (s)
Treating diabetes	Low glycemic index starch
Helps in anemia by increasing hemoglobin	High iron and zinc concentration
Alleviates constipation	High fiber
Anti-cancer inhibiting tumor formation	Vanillin (antioxidant)
Bone growth, development, and repair	High phosphorus (P)
Prevents stomach ulcers, preventing formation of excess acidity	Gastroprotective effect of antioxidant in millet
Helps alleviating high blood pressure and heart stress	Polyphenols, lignin, high magnesium (Mg)
Reduces respiratory problems in asthma patients and reduce migraine attacks	High magnesium (Mg)
Reduces the overall consumption of food	High fiber content
Reduces the risk of gall bladder stone and production of excessive bile juice	High fiber content
Highly digestible and low allergic response (Reduces allergy in celiac patients)	Gluten-free starch

Note: The data was adapted from Malik (2015)

reducing powers (Saleh et al. 2013). Several antioxidants present in millets neutralize free radicals (reactive oxygen species) and boost overall immunity and human health. The free radicals can lead to oxidative stress, which in turn are associated with conditions like cancer, arthritis, respiratory, and compromised immunity. Thus, consumption of millets can improve human health by scavenging free radicals from the liver and kidney.

4.2 Antidiabetic

Diabetes mellitus causes several body disorders, characterized by high blood sugar levels with an imbalance in protein, carbohydrate, and lipid metabolism. Consumption of whole millet grains has been clinically proven to lower diabetic effects (Saleh et al. 2013). Rao et al. (2017) reported that regular consumption of millets helps prevent diabetes due to the higher level of magnesium in these small grains, which in turn enhances the efficiency of insulin and glucose receptors. Furthermore, millet-based diets have been reported for their antioxidant potential, nerve growth factor production, and wound healing properties (Rao et al. 2017).

4.3 Cardiovascular Diseases

Millets are abundant in magnesium, potassium, lignans, antioxidants, and fibers, which can reduce the risk of heart strokes, high blood pressure, low-density cholesterol, and atherosclerosis. Rao et al. (2017) reported that the addition of barnyard,

proso, and finger millets in the diet increased the blood plasma levels of adiponectin, high-density lipoprotein (HDL), and reduced triglycerides in hyperlipidemia. Therefore, millets, including proso, finger, barnyard, foxtail, and pearl millet, can significantly lower hyperlipidemia, hyperglycemia, triglycerides, and non-HDL and reduce the risk of cardiovascular diseases (Rao et al. 2017).

4.4 Anticancer

Cancer is characterized by changes at the cellular level due to uncontrolled cell growth and division. Millets generally have high levels of phenolic compounds such as tannins, fibers, and other phytonutrients than other cereals. These compounds improve health status and lower the risks of colon, breast, and esophageal cancers (Saleh et al. 2013). It is recommended that a daily ingestion of 30 g of fiber may reduce the risk of breast cancer chances by 50% (Rao et al. 2017).

4.5 Antiaging

Millets, including proso millet, are nutrient-rich functional food, enriched with high levels of micronutrients, vitamins, minerals, phenolic compounds, fibers, and antioxidants. These compounds are helpful in inhibiting glycation and cross-linking of collagen which is responsible for aging (Saleh et al. 2013).

4.6 Gluten Sensitivity

Celiac disease and gluten sensitivity are caused by gluten in organisms that are genetically susceptible to these conditions. The disease affects 0.6–1% of the population worldwide. Other food grains are amaranth, buckwheat, quinoa, corn, and sorghum functional and nutraceutical food. Millets have a lot of potential in the production of gluten-free food and beverages targeting people suffering from gluten sensitivity and celiac disease (Saleh et al. 2013).

5 Genetic Variation of Seed Components

Genetic variation is crucial for the adaptability and survivability of any species. From a breeding perspective, genetic variation ensures the selection of superior individuals (plants or animals) with improved characteristics from a population. A very limited number of studies have been conducted to assess the variation in seed composition in proso millet. Ravindran (1992) studied the amino acid composition and a few other traits of six proso millet varieties. He reported small differences in amino acid content among the varieties (Ravindran 1992). Kalinova and Moudry

(2006) assessed the crude protein and amino acid content in seven proso millet varieties. In contrast to Ravindran's findings, they observed significant varietal differences in seed content of the amino acid lysine, valine, isoleucine, phenylalanine, aspartate, glycine, tyrosine, histidine, and arginine (Kalinova and Moudry 2006). In another study, a varietal difference in phenolic compounds and antioxidant capacity in whole millet grains was reported (Chandrasekara and Shahidi 2011). Vetriventhan and Upadhyaya (2018) studied 200 proso millet accessions to assess the diversity of grain nutritional traits including seed protein, calcium (Ca), iron (Fe), and zinc (Zn). Large variations for all four grain-nutritional traits were observed in their study. More research on the genetic variation of seed components, including secondary metabolites, needs to be conducted for generating resources for genome-wide association studies (GWAS) and developing varieties with improved nutrient profiles (Vetriventhan and Upadhyaya 2018).

6 Breeding for Grain Quality Improvement

Plant breeding is the process of creating and selecting superior genotypes with enhanced yield, abiotic and biotic stress tolerance, and improved nutritional quality. Breeding programs dedicated to developing genetically improved varieties of proso millet were established in several countries. China, the USA, India, Russia, and Kenya are notable countries that have been instrumental in continuing efforts for the genetic improvement of proso millet (Gomashe 2017). Considerable progress has been made in generating high-yielding varieties with desirable maturity, seed color, plant height, and other yield-related morpho-agronomic traits. However, efforts to develop superior proso millet cultivars enriched with nutrients and health-promoting metabolites are very limited. With the growing demand for proso millet in the ever-expanding human healthy food market, it is imperative that breeding programs focused on improving nutrient and metabolite profiles of the high-yielding varieties are established to promote this crop as a functional food. Common breeding approaches and a couple of biotechnological tools to complement conventional breeding in accelerating variety development efforts are discussed further down.

6.1 Conventional and Wide Hybridization

The majority of proso millet varieties of North America were developed through conventional breeding methods (crossing followed by selection of progenies) using a limited pool of parental lines through conventional hybridization. The resulting narrow genetic base may lead to diminished performance of the varieties in terms of resistance to abiotic and biotic stresses over time. Besides, the variation of seed quality traits is not very large, which makes genetic improvement of seed composition through conventional breeding approaches even tougher. Nevertheless, the conventional method was still useful in developing specialty proso millet varieties

for human food uses. One such example was 'Plateau', the first and only waxy proso millet variety in North America (Santra et al. 2015). The variety was developed via artificial pollination between PI 436626 (Lung Shu 18), a Chinese waxy accession, and 'Huntsman', a high-yielding US variety. Similar market-driven strategies can be adopted to develop new high-yielding cultivars enriched with essential nutrients and healthy metabolites to market them as nutraceuticals.

The gene pool of wild relatives of proso millet can be a large repository of novel traits or genes that can be transferred via introgression into the existing cultivars (Santra et al. 2019). Wide hybridization is a promising prospect for improving grain nutritional quality by introducing genes from the wild relatives such as P. *sumatrense* and *P. repens* to improve the metabolic profile of the cultivars.

6.2 Mutation Breeding

Mutation breeding involves deliberate induction of mutation by artificial mutagenesis and breeding for genetic variations that do not exist in the gene pool. Mutation-inducing agents (mutagens) can be physical, ionizing (X-rays and γ-rays), and non-ionizing radiations (ultraviolet rays) or chemical agents such as ethylmethane sulfonate (EMS), methyl methane sulfonate (MMS), diethylsulfate (DES), and nitrosoguanidine (NS). EMS is the most widely used chemical mutagen due to its effectiveness and ease of use. Proso millet seeds treated with 0.2% EMS was found to improve panicle filling, grain shattering tolerance, and grain yield in the M2 progeny (Singode et al. 2018). Recently, Francis et al. (2022) used different doses of gamma radiation to develop phenotypic mutations for various traits, including plant height, panicle shape, compactness, seed color, and lodging resistance. They were able to isolate eight high-yielding mutant families with significantly improved grain yield (GY) and fodder yield (FY) (Francis et al. 2022). More research on artificial mutagenesis is absolutely required in proso millet. Mutation breeding is a viable technique to overcome poor combining ability and narrow genetic based for improving grain nutrient and metabolite profile (Santra et al. 2019).

6.3 Doubled Haploids and Transgenics

Doubled haploid (DH) and transgenic technologies are two commonly used tissue culture-based biotechnology tools for the genetic improvement of crops. The DH technology produces 100% homozygous genotypes using haploid gametes (pollen or egg). Transgenic technology uses genetic engineering to insert one or two specific genes from other plants. Conventional plant breeding via pure-line selection normally takes 11–13 years to develop a new variety as it usually involves 5–6 generations of inbreeding to generate homogeneous breeding lines with >98% homozygosity before multi-year yield assessment at multiple sites. The desirable 100% homozygosity can be achieved in a much shorter time by implementing doubled haploid technology (Khound et al. 2013; Santra et al. 2019). Only a few

reports on successful tissue regeneration have been reported in proso millet. In 1982, Heyser and Nobors reported successful callus induction and shoot regeneration from a variety of explants by manipulating auxin levels in Linsmaier and Skoog (L and S) medium (Heyser and Nabors 1982). Bobkov and Suvorova (2012) studied the efficiency of anther culture technique in proso millet for embryogenic callus induction and regeneration. Heat (32 °C) and cold (4 °C) were observed to successfully induce callus formation and regeneration of the explants (Bobkov and Suvorova 2012). For several decades, genetic engineering has been used to introduce foreign genes expressing novel traits into a host genome or to knock out genes with deleterious effects. Genetic improvement of crops through genetic engineering relies on the development of an efficient regeneration method and a robust transformation system (Kumar et al. 2016). There has been no report on the generation of transgenic proso millet via genetic transformation so far. The DH and genetic engineering technologies can be used for developing proso millet varieties enriched with health-promoting bioactive compounds and no or minimum antinutrients. Therefore, more studies on efficient tissue culture and transformation systems are warranted for facilitating genetic engineering-assisted genetic improvement of proso millet.

7 Omics for Improving Grain Nutritional Quality

Biology works as a "system" rather than as an individual part separately. All the parts of biology work together in harmony to manifest overall biological functions. Many biochemical compounds are involved in this intrinsic process. Similarly, all the nutrients and antinutrients in food work together in harmony to manifest their effect on human health and nutrition. Therefore, it is very important to study proso millet seed nutrients quality and quantity, genes, biosynthesis pathways, regulation, and interactions together rather than separately. That means all aspects of proso millet seed nutraceuticals must be studied following the "omics" approach. The term "neutrocetomics" is used to indicate the combined study of nutraceuticals and omics. Proso millet being a minor crop, its "omics" areas of research are not as rich as other major crops such as corn, wheat, rice, and soybean. However, in recent years a lot of omics research has been done in proso millet, and more is being conducted due to its increasing importance for human health. Details of proso millet "omics" research were published in a recent comprehensive review paper (Khound and Santra 2020). The following section summarizes various fields of "omics" in proso millet and their relevance to its "nutraceutomics" (Fig. 3).

7.1 Genomics

Genomics is the branch of omics that tackles large-scale studies on the structural and functional aspects of the genome to understand the genetic and molecular underpinning of various biological processes. Initial genomic studies were primarily focused on genome size and physical as well as genetic mapping of organisms. Next-

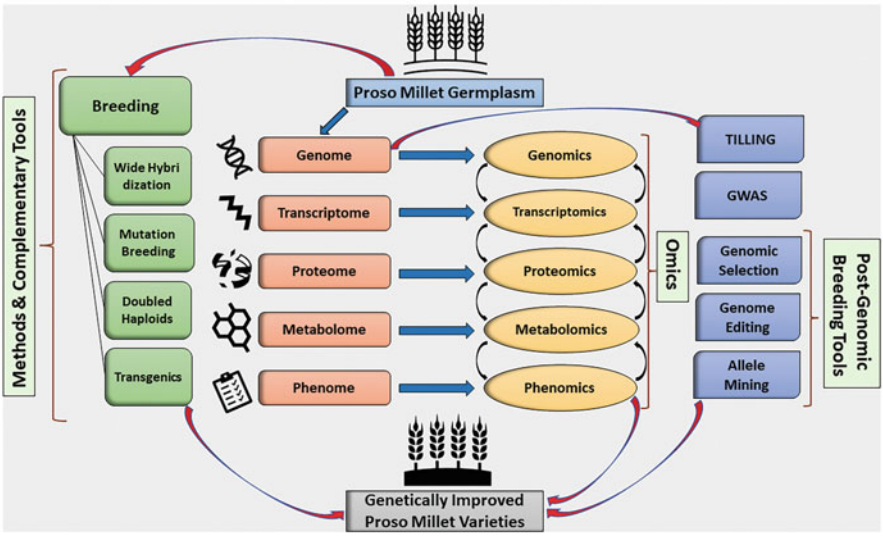


Fig. 3 Schematic representation of the applications of traditional breeding methods and tools, as well as omics-based and post-genomic breeding strategies to the genetic improvement of proso millet

generation sequencing (NGS) marked a new era of genomics by enabling researchers to sequence, assemble, and analyze the genomes of important crops at lower costs and in a shorter time (Khound and Santra 2020).

7.1.1 Molecular Markers

Molecular markers are important genomic resources that have been widely used in genetic mapping, genetic diversity, and taxonomic and population genetic studies of crops. A limited number of molecular markers were developed in proso millet as this crop is grossly under-researched. The molecular markers identified in proso millet so far include amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), inter-simple sequence repeat (ISSR), cleaved amplified polymorphic DNA (CAP), simple sequence repeat (SSR), and single nucleotide polymorphism (SNP) markers. These molecular markers were primarily used for evaluating the genetic diversity of the proso millet germplasm (Santra et al. 2019).

SSRs are preferable for genetic studies as they are abundant, evenly distributed, multi-allelic, codominant, highly polymorphic, easy to score, and highly reproducible (Khound and Santra 2020). The earliest SSR-based genetic studies were conducted using SSRs from other species such as rice, wheat, oat, barley, and switchgrass. The first proso millet-specific SSRs were identified by Cho and coworkers (Cho et al. 2010). Twenty-five polymorphic SSRs were developed from the genomic DNA of 50 diverse proso millet genotypes. These SSRs were used to evaluate the genetic diversity of Chinese proso millet accessions (Liu et al. 2016).

SNP markers are ideal for genetic studies, genetic map constriction, and genomic selection as they occur abundantly in the genome (Khound and Santra 2020). These markers are especially well suited for constructing linkage maps as they are cost-effective, high-throughput, and more efficient compared to other DNA markers (Rajput et al. 2016). Several reports on SNP identification in proso millet have been published. Rajput et al. identified 833 high-quality biallelic genotype-by-sequencing (GBS)-SNPs from genotyping of 93 recombinant inbred lines (RILs). These GBS-SNPs were used for constructing the first proso millet linkage map (details later in the quantitative trait locus (QTL) mapping section) (Rajput et al. 2016). Wang and her coworkers evaluated the allelic diversity of two waxy genes, viz., Wx-L and Wx-S, in 132 proso millet accessions from 12 provinces of China. In their study, Wang et al. (2018) identified six SNPs at the *Wx-L* locus after sequencing the PCR amplicons of all the accessions. Johnson and his team used GBS to generate 1882 filtered SNPs to conduct genome-wide population genetic studies in proso millet. The SNP markers were found to accurately represent genetic variation within the population (Johnson et al. 2019). In 2021, Boukail et al. (2021) developed 2412 SNPs from 88 proso millet accessions with RAD-seq. The identified SNPs were used for conducting GWAS to detect marker-trait associations for several agronomic and seed morphology traits (more details are covered in Sect. 8) (Boukail et al. 2021). In another recently published report, 126,822 filtered SNPs were identified using specific-locus amplified fragment sequencing (SLAF-seq) of 106 accessions (Li et al. 2021). More recently, Khound et al. (2022) developed 972,863 high-quality biallelic SNPs from low-pass genome sequencing of 85diverse proso millet genotypes of the USDA gene bank. They employed those SNPs to study the population structure and phylogenetic relationships among the genotypes (Khound et al. 2022).

7.1.2 QTL Mapping

Quantitative trait locus (QTL) mapping is routinely used in many crops for detecting quantitative trait loci (QTLs) associated with key traits due to the abundance of versatile DNA markers and statistical models and methods. QTLs for a wide array of traits have been reported in many major and minor crops including wheat, rice, maize, sorghum, millet, amaranth, quinoa, oat, and rye (Yabe and Iwata 2020). Among millets, pearl millet and foxtail millet have the major share of published reports on QTL mapping. In proso millet, very little research has been done on genetic mapping of QTLs compared to major crops such as rice, wheat, maize, soybean, and sorghum. This is possibly because of the minor crop status and inadequate resources available to proso millet geneticists. The first-ever genetic linkage map of proso millet was published in 2016 by Rajput et al. They used 833 GBS-SNPs and 93 RILs. Several QTLs and linked SNP markers for a few morpho-agronomic traits were identified (Rajput et al. 2016). However, no QTLs for proso millet seed components such as starch, protein, minerals, vitamins, and other bioactive compounds have been identified in proso millet to date.

7.1.3 Whole-Genome Assembly

Developing a whole-genome assembly (WGA) is a crucial step in genome-wide molecular studies as this creates opportunities for exploring complex molecular structures and functions in an organism. Information on genome assembly is an invaluable resource for genomic-assisted breeding for crop improvement. Proso millet was the fourth millet to have a published whole-genome assembly after foxtail millet, pearl millet, and finger millet (Khound and Santra 2020). In 2019, Zou et al. developed the complete proso millet genome sequence using various NGS tools. They also identified 55,930 protein-coding and 339 microRNA genes (Zou et al. 2019). In the same year, Shi et al. reported a near-complete assembly of the proso millet genome. The authors generated 18 super scaffolds covering approximately 96% of the estimated genome. Moreover, they were able to annotate 63,671 protein-coding genes in the proso millet genome (Shi et al. 2019).

7.1.4 TILLING

Another modern approach that could be applied for improving grain nutritional quality of proso millet is the TILLING (targeted induced local lesions in genomes) method. This is an advanced method of detecting beneficial mutations or alleles that allows rapid identification of induced gene mutations within a mutagenized population via heteroduplex analysis. The initial step of TILLING involves creating a mutagenized population, which is subsequently assessed for detecting useful gene mutations that could be linked to important phenotypes. The mutations within the mutagenized population are usually detected by one of the three methods, viz., LI-COR method (uses CEL 1 enzyme), high-resolution melting (HRM) method, and NGS method. TILLING has been used for detecting useful mutations in a wide array of crops including wheat, rice, sorghum, and maize (Irshad et al. 2020). Tilling has been successfully employed to identify mutations in genes associated with starch synthesis, plant architecture, and disease resistance in several crop species. Another related method, namely, EcoTILLING, is used to detect naturally occurring polymorphisms within a population. The polymorphisms identified could be used to study the phylogenetic diversity within the population. This technique can also be used to detect novel allelic variations that can be exploited for genetic improvement. EcoTILLING has been utilized in different crops including Arabidopsis, chickpea (seed weight), wheat (disease susceptibility), rice (drought tolerance, starch synthesis), and soybean (seed protein) (Irshad et al. 2020).

7.1.5 Allele Mining

A large number of beneficial alleles are present in plant genetic resources, which remain unutilized as they were abandoned during evolution and domestication. Introgression of these natural genetic variations has the potential to improve the performance of available cultivars. The recent surge of genomic data accelerated the discovery and annotation of novel genes and loci linked to essential agronomic traits in many crop species. The idea of identifying alleles in the annotated genes led to the concept of allele mining in plants. Allele mining refers to the approach of dissecting

naturally occurring variations at candidate genes or loci associated with major agronomic traits with the help of modern genomic tools (Kumar et al. 2010). Two primary approaches for identifying alleles are EcoTILLING and sequence-based allele mining. As discussed in the previous section, EcoTILLING is used to identify naturally occurring allelic variants present in the primary and secondary gene pools. In a sequence-based approach, alleles of diverse genotypes are amplified by PCR and subsequently sequenced. Allele mining was used in identifying alleles for important seed quality traits including *Amy32b(*α-amylase gene) and *Gpc-B1* (gene for grain protein content) in barley, *wx* locus (waxy gene) in rice, and *Wx-A1* (waxy gene) in wheat (Kumar et al. 2010). The prospect of using allele mining for detecting novel alleles of genes is relatively unexplored in millets. The availability of whole-genome assemblies, high-quality molecular markers, and the gradually reducing cost of sequencing is expected to result in more studies for identifying novel genes and loci. That would encourage more efforts for the mining of candidate genes for some key agronomic and grain quality traits leading to the detection of useful alleles for application in molecular breeding.

7.2 Transcriptomics

A handful of reports on transcriptome studies investigating biological processes have been published in proso millet. Yue et al. (2016) used Illumina sequence reads from two proso millet accessions, viz., Yumi No.2 and Yumi No.3, to assemble a proso genome transcriptome. From the 113,643 unigenes assembled, 62,543 contigs were assigned to 315 gene ontology (GO) categories. Additionally, 15,514 unigenes could be mapped into 202 Kyoto Encyclopedia of Genes and Genomes (KEGG) clusters, and 51,020 unigenes were mapped to 25 clusters of orthologous groups (COG) categories. The most represented KEGG pathways included metabolic pathways (25.65%), biosynthesis of secondary metabolites (10.71%), and biosynthesis of amino acids (3.57%) (Yue et al. 2016). In 2017, Hou et al. developed a transcriptome assembly using Illumina sequence reads from a single genotype of proso millet named Neimenggu-Y1. They were able to identify 25,341 unigenes out of which 5170 (20.4%) could be mapped to 146 KEGG pathways. GO annotation with 2936 tissue-specific genes resulted in three subcategories- biological processes, molecular functions, and cellular components (Hou et al. 2017). Zhang et al. (2019) published a comparative analysis of transcriptome associated with drought tolerance in two genotypes, viz., Neimi 5 and Jinshu 6. Drought stress resulted in 833 and 2166 DEGs in Jinshu 6 and Neimi 5, respectively. Some of the DEGs could be mapped to some key KEGG pathways including carbon metabolism, phenylpropanoid biosynthesis, and amino acid biosynthesis (Zhang et al. 2019). The unigenes that were mapped to the key KEGG pathways in the above reports should be explored further to identify their roles in those pathways. This information could be useful in targeting specific genes for enriching proso millet with useful nutraceuticals or inhibiting the biosynthesis of antinutrients.

7.3 Proteomics

There is an abundance of reports on proteomics research for the major crops including rice, corn, soybean, wheat, and potato. However, like other omics disciplines, the number of published reports pertaining to proteomics in proso millet and other millet is very meager (Khound and Santra 2020). Perhaps the first proteomics analysis in proso millet was done on a 2500-year-old starch food in China. Comparative proteomics analysis between the ancient sour dough bread and a few other reference cereals identified proso millet and barley as the ingredients of the ancient bread (Shevchenko et al. 2014). In 2017, Roy and co-workers reported the seed protein analysis of four Korean proso millet varieties using two-dimensional (2-D) electrophoresis and mass fingerprinting to map the seed proteins and determine their functional properties. They detected 1152 differentially expressed proteins, out of which 26 reproducible proteins were further analyzed using matrix-assisted laser desorption/ionization time-of-flight/time-of-flight mass spectrometry (MALDI-TOF-TOF/MS). Two of the 26 proteins were found to be upregulated in all the cultivars, while 13 were upregulated and 11 were downregulated in 2 proso millet cultivars. The authors opined that the differential expression of the proteins in the four proso millet cultivars was possibly variety-specific (Roy et al. 2017).

7.4 Metabolomics

There are only a couple of available reports on the metabolomics of proso millet. The first published study on proso millet metabolomics used gas chromatography-time-of-flight mass spectrometry (GC-TOFMS) to evaluate gain quality. Kim et al. (2013) studied the primary metabolites and phenolic acids of the matured grains of three Korean proso millet varieties, viz., 'Joongback', 'Joongjuk', and 'Hwangguem'. They were able to identify 48 metabolites from the grains, which included 43 primary metabolites and 5 phenolic acids. The mature grains of the variety 'Joongjuk' contained significantly higher levels of phenolic acids than the other two varieties. This makes this variety a suitable candidate for further evaluations and genetic improvement as a nutraceutical (Kim et al. 2013).

In a relatively recent publication, 172 metabolites and 3 cooking quality traits were compared between conventionally and organically grown seeds of two proso millet varieties. There was no difference in the metabolite profiles between the conventionally and organically grown gains, except in the levels of some carbohydrates such as glucose and fructose, which were higher in the organically grown grain. The variations observed in the metabolite content could be primarily attributed to the variety (Liang et al. 2018a). These findings emphasize the importance of variety selection for developing proso millet varieties for nutraceutical use.

7.5 Phenomics

A phenome can be described as the complete set of phenotypic traits expressed by a cell, tissue, or organism, and the field of studies on the phenome is known as phenomics. Establishing the relationship between the genotypes and phenotypes is a major breeding objective of any crop improvement program. Traditional phenotyping approaches are usually laborious, time-consuming, expensive, and mostly destructive (Santra et al. 2019). It is important to determine more phenotype-to-genotype relationships to develop reliable predictive models for predicting a full array of phenotypes of a genotype (Gustin and Settles 2015). In recent years, high-throughput phenotyping techniques and tools including high-resolution imaging, spectroscopy, robotics, and powerful algorithms have been developed to push plant phenotyping to the next level. Different high-throughput phenotyping approaches have been used to study a variety of phenotypic traits such as plant growth, biomass, leaf morphology, maturity, and nutrient status (Santra et al. 2019). To the best of our knowledge, the first-ever report on high-throughput phenotyping in proso millet was reported by Zhao et al. who used UAV-based imaging for heading percentage detection in proso millet (Zhao et al. 2022).

A more niche area of phenomics, for example, "seed phenomics," could be developed in proso millet to specifically study various seed characteristics (Gustin and Settles 2015). This will require the integration of various imaging technologies, spectroscopy, and multiple omics such as genomics, transcriptomics, proteomics, and metabolomics. Various imaging techniques and rapid, nondestructive spectroscopic techniques, such as near-infrared (NIR) spectroscopy and nuclear magnetic resonance (NMR) spectroscopy, are routinely used to determine various seed traits such as shape, size, color, and chemical composition. To the best of our knowledge, there is no report or ongoing research on proso millet seed phenome, which is very important for genetic improvement of nutraceuticals values in proso millet.

8 Post-Genomic Approaches for Improving Seed Nutritional Quality

The following sections summarize the current status of proso millet "omics" resources available to the scientists working in the field of proso millet nutraceutomics. The following sections address post-genomic omics approaches for genetic manipulation of proso millet seed nutraceuticals.

8.1 Genome-Wide Association Studies (GWAS)

Genome-wide association study (GWAS) or association mapping (AM) involves the detection of an association between DNA marker(s) and a trait of interest based on the principle of linkage disequilibrium (LD). This is achieved through

large-scale genotyping of germplasm panels or breeding populations exhibiting contrasting phenotypes across different environments. Several GWAS have been conducted in millets to uncover the marker-trait association for some important traits. However, only a few reports of GWAS for seed nutritional traits are available. The only available report on GWAS in proso millet to date was published in 2021 by Boukail et al. They used a global collection of 88 varieties and landraces to identify marker trait associations (MTAs) for seed morphological traits. They identified 2412 high-quality SNPs using restriction site-associated DNA sequencing (RAD-seq). These SNPs were used for GWAS for seed traits, such as seed length (SL), seed width (SW), seed perimeter (SP), and seed color (RGB) as well as agronomical traits. They identified MTAs for the seed and agronomic traits (Boukail et al. 2021). The SNPs that were found to have a strong association with the agronomic and seed traits could be strong candidates for marker-assisted selection (MAS) in proso millet breeding programs. Having said that, no report on GWAS for nutraceutical traits is available in proso millet yet. This warrants the initiation of genome-scale studies to identify MTAs for some key seed quality traits to accelerate the development of proso millet varieties with impeccable agronomic qualities and healthy seed components.

8.2 Genomic Selection (GS)

Genomic selection (GS) is the breeding approach of using genome-wide high-density markers to facilitate rapid selection of suitable candidates for breeding (Srivastava et al. 2020). GS is still in the nascent stage in millets as very few GS studies have been conducted so far due to the limited availability of genomic resources. Varshney et al. (2017) utilized whole-genome resequencing (WGRS) data for performing GS to predict the grain yield of pearl millet under four different stress scenarios across environments. They used to analyze the grain yield of 64 hybrids with 302,110 SNPs to identify promising hybrid combinations for hybrid production (Varshney et al. 2017). In another study, Liang and his coauthors (2018b) evaluated four genomic selection schemes using two genotyping strategies, namely, RAD-seq and tunable genotyping by sequencing (tGBS) in pearl millet. The authors observed that for traits with significant mid-parent heterosis, the inbred phenotypic data moderately improved genomic predication of the hybrid genomic estimated breeding values when the trait values of the inbred and hybrid lines were scored relative to the mean trait values of the corresponding populations (Liang et al. 2018b). Similar GS studies can be conducted in proso millet for major morpho-agronomic and seed quality traits to accelerate genetic improvement of breeding populations in a time-efficient and cost-effective manner.

8.3 Genome Editing

Genome editing is a relatively newer technique adopted by plant breeders to develop new and improved varieties of crops. Rather than introducing transgene(s)

randomly into the host genome, genome editing used sequence-specific nucleases (SSNs) to induce targeted and precise nucleotide sequence changes to the genome (Santra et al. 2019). Genome editing tools have been successfully used to introduce genes into major cereals including rice, wheat, and maize (Ceasar 2022). The most commonly used genome editing tools are clustered regularly interspersed short palindromic repeats-CRISPR-associated nucleases (CRISPR-Cas) system, transcription activator-like effector nucleases (TALENs), and zinc-finger nucleases (ZFNs) (Santra et al. 2019; Ceasar 2022).

There are numerous reports on genome editing in major cereals, especially rice, wheat, maize, and barley. However, genome editing is still a relatively unexplored territory in millets. One example of the successful use of genome editing tools in improving grain quality is the knock-out of three key genes associated with phytic acid (PA) biosynthesis in maize. PA naturally occurs in the grains of many cereal crops, including proso millet. It is considered an antinutrient as it is largely indigestible and may cause environmental pollution. TALEN and CRISPR-Cas9 systems were used to induce mutations in the genes, viz., *ZmIPK*, *ZmIPK1A*, and *ZmMRP4*, encoding enzymes that catalyze three steps in PA biosynthesis (Liang et al. 2014). There are only two reports on successful CRISPR-Cas9-mediated genome modification in foxtail millet (Ceasar 2022). Similar strategies can be implemented in proso millet for improving grain quality by reducing the levels of antinutrients and enriching with health-promoting compounds such as carotenoids. However, reliable, reproducible, and robust micropropagation and transformation systems need to be developed for this crop in order to accomplish this.

9 Conclusion and Future Prospects

Compared to the major crops such as rice, wheat, corn, and soybean, there is a dearth of genomic resources available for the genetic improvement of proso millet. Especially, little to no progress has been made in exploring genes or QTLs linked to grain nutraceutical traits. Therefore, there is a need for devising strategies for developing and harnessing genomic resources in the identification of genes and QTLs associated with seed components in proso millet. This will facilitate omics-assisted breeding of this ancient crop thereby enabling rapid and precise genetic improvement for various agronomic and seed quality traits. Therefore, the global millet researchers must work together to take advantage of such technological advancement in the "omics" research to advance the field of "nutraceutomics" in proso millet. Interdisciplinary research, extension, and promotion are essential for using proso millet in the human food market.

The future prospects of proso millet along with other climate-resilient millets are vast for the food and nutritional security of the global population in the current century and future. The global climate is deteriorating exponentially with the acute pressure of population increase and farmland reduction. All seven millets have great

potential to address the mammoth global challenge. A strong collaboration among the millet geneticists and breeders across the globe will be mandatory for the successful implementation of the "omics" in proso millet "nutraceutomics" research and applications for human health and nutritional security in the changing climate, especially in the climate-fragile countries in the world.

Conflict of Interest The authors declare that they have no conflict of interest.

Funding Acknowledgment The project was supported by "Research State Aided" internal funds: 21-6243-1001.

References

Akharume F, Santra D, Adedeji A (2020) Physicochemical and functional properties of proso millet storage protein fractions. Food Hydrocoll 108:105497. https://doi.org/10.1016/j.foodhyd.2019.105497

Anderson RL, Bowman RA, Nielsen DC et al (1999) Alternative crop rotations for the central Great Plains. J Prod Agric 12:95–99. https://doi.org/10.2134/jpa1999.0095

Argos P, Pedersen K, Marks MD, Larkins BA (1982) A structural model for maize zein proteins. J Biol Chem 257:9984–9990. https://doi.org/10.1016/S0021-9258(18)33974-7

Asharani VT, Jayadeep A, Malleshi NG (2010) Natural antioxidants in edible flours of selected small millets. Int J Food Prop 13:41–50. https://doi.org/10.1080/10942910802163105

Baltensperger DD (2002) Progress with proso, pearl and other millets. In: Janick J, Whipkey A (eds) Trends in new crops and new uses. ASHS Press, Alexandria, VA, pp 100–103

Bangar SP, Ashogbon AO, Dhull SB et al (2021) Proso-millet starch: properties, functionality, and applications. Int J Biol Macromol 190:960–968. https://doi.org/10.1016/j.ijbiomac.2021.09.064

Bobkov S, Suvorova G (2012) Temperature stress in anther culture of proso millet (*Panicum miliaceum* L.). In: Yan C, Baili F (eds) Proceedings of the 1st international broomcorn millet symposium: advances in broomcorn millet research, pp 82–88

Boukail S, Macharia M, Miculan M et al (2021) Genome wide association study of agronomic and seed traits in a world collection of proso millet (*Panicum miliaceum* L.). BMC Plant Biol 21:330. https://doi.org/10.1186/s12870-021-03111-5

Ceasar A (2022) Genome-editing in millets: current knowledge and future perspectives. Mol Biol Rep 49:773–781. https://doi.org/10.1007/s11033-021-06975-w

Chandrasekara A, Shahidi F (2011) Inhibitory activities of soluble and bound millet seed phenolics on free radicals and reactive oxygen species. J Agric Food Chem 59:428–436. https://doi.org/10.1021/jf103896z

Cho Y-I, Chung J-W, Lee G-A et al (2010) Development and characterization of twenty-five new polymorphic microsatellite markers in proso millet (*Panicum miliaceum* L.). Genes Genom 32:267–273. https://doi.org/10.1007/s13258-010-0007-8

Das S, Khound R, Santra M, Santra D (2019) Beyond bird feed: proso millet for human health and environment. Agriculture 9:64. https://doi.org/10.3390/agriculture9030064

Demirbas A (2005) Glucan and mineral nutrient contents of cereals grown in Turkey. Food Chem 90:773–777. https://doi.org/10.1016/j.foodchem.2004.06.003

Dendy DAV (1995) Sorghum and Millets: Chemistry and Technology. AACC, St. Paul, Minnesota, 406 pp.

Francis N, Rajasekaran R, Krishnamoorthy I et al (2022) Gamma irradiation to induce beneficial mutants in proso millet (*Panicum miliaceum* L.): an underutilized food crop. Int J Radiat Biol 98:1277–1288. https://doi.org/10.1080/09553002.2022.2024292

Geervani P, Eggum BO (1989) Nutrient composition and protein quality of minor millets. Plant Foods Hum Nutr 39:201–208. https://doi.org/10.1007/BF01091900

Gomashe SS (2017) Proso millet, *Panicum miliaceum* (L.): genetic improvement and research needs. In: Patil JV (ed) Millets and sorghum. John Wiley & Sons, Ltd, Chichester, pp 150–169

Gulati P, Li A, Holding D et al (2017) Heating reduces proso millet protein digestibility via formation of hydrophobic aggregates. J Agric Food Chem 65:1952–1959. https://doi.org/10.1021/acs.jafc.6b05574

Gulati P, Jia S, Li A et al (2018) In vitro pepsin digestibility of cooked proso millet (*Panicum miliaceum* L.) and related species from around the world. J Agric Food Chem 66:7156–7164. https://doi.org/10.1021/acs.jafc.8b02315

Gustin JL, Settles AM (2015) Seed phenomics. In: Fritsche-Neto R, Borém A (eds) Phenomics. Springer International Publishing, Cham, pp 67–82

Habiyaremye C, Matanguihan JB, D'Alpoim Guedes J et al (2017) Proso millet (*Panicum miliaceum* L.) and its potential for cultivation in the Pacific Northwest, U.S.: a review. Front Plant Sci 7. https://doi.org/10.3389/fpls.2016.01961

Heyser JW, Nabors MW (1982) Regeneration of proso millet from embryogenic Calli derived from various plant Parts1. Crop Sci 22:1070–1074. https://doi.org/10.2135/cropsci1982.0011183X002200050043x

Hou S, Sun Z, Li Y et al (2017) Transcriptomic analysis, genic SSR development, and genetic diversity of proso millet (*Panicum miliaceum*; Poaceae). Appl Plant Sci 5:1600137. https://doi.org/10.3732/apps.1600137

Hunt HV, Badakshi F, Romanova O et al (2014) Reticulate evolution in Panicum (Poaceae): the origin of tetraploid broomcorn millet, *P miliaceum*. J Exp Bot 65:3165–3175. https://doi.org/10.1093/jxb/eru161

Irshad A, Guo H, Zhang S, Liu L (2020) TILLING in cereal crops for allele expansion and mutation detection by using modern sequencing technologies. Agronomy 10:405. https://doi.org/10.3390/agronomy10030405

Johnson M, Deshpande S, Vetriventhan M et al (2019) Genome-wide population structure analyses of three minor millets: Kodo millet, little millet, and proso millet. Plant Genome 12:190021. https://doi.org/10.3835/plantgenome2019.03.0021

Jones RW, Beckwith AC, Khoo U, Inglett GE (1970) Protein composition of proso millet. J Agric Food Chem 18:37–39. https://doi.org/10.1021/jf60167a003

Kalinová J (2007) Nutritionally important components of proso millet (*Panicum miliaceum* L.). FoodReview 1:91–100

Kalinova J, Moudry J (2006) Content and quality of protein in proso millet (*Panicum miliaceum* L.) varieties. Plant Foods Hum Nutr 61:43–47. https://doi.org/10.1007/s11130-006-0013-9

Kalra EK (2003) Nutraceutical-definition and introduction. AAPS PharmSci 5:27–28. https://doi.org/10.1208/ps050325

Khound R, Santra DK (2020) Omics for proso millet genetic improvement. Nucleus 63:241–247. https://doi.org/10.1007/s13237-020-00339-8

Khound R, Santra M, Baenziger PS, Santra DK (2013) Effect of cold-mediated pretreatment on microspore culture in winter and spring wheat. Am J Plant Sci 04:2259–2264. https://doi.org/10.4236/ajps.2013.411278

Khound R, Sun G, Mural RV et al (2022) SNP discovery in proso millet (*Panicum miliaceum* L.) using low-pass genome sequencing. Plant Direct 6(9):e447. https://doi.org/10.1002/pld3.447

Kim JK, Park S-Y, Yeo Y et al (2013) Metabolic profiling of millet (*Panicum miliaceum)* using gas chromatography–time-of-flight mass spectrometry (GC-TOFMS) for quality assessment. Plant Omics J 6:73–78

Kumar GR, Sakthivel K, Sundaram RM et al (2010) Allele mining in crops: prospects and potentials. Biotechnol Adv 28:451–461. https://doi.org/10.1016/j.biotechadv.2010.02.007

Kumar A, Metwal M, Kaur S et al (2016) Nutraceutical value of finger millet [*Eleusine coracana* (L.) Gaertn.], and their improvement using omics approaches. Front Plant Sci 7:934. https://doi.org/10.3389/fpls.2016.00934

Kumari SK, Thayumanavan B (1998) Characterization of starches of proso, foxtail, barnyard, kodo, and little millets. Plant Foods Hum Nutr 53:47–56

Ladabaum U, Mannalithara A, Myer PA et al (2014) Obesity, abdominal obesity, physical activity, and caloric intake in US adults: 1988 to 2010. Am J Med 127(8):717–727.

Li C, Liu M, Sun F et al (2021) Genetic divergence and population structure in weedy and cultivated broomcorn millets (*Panicum miliaceum* L.) revealed by specific-locus amplified fragment sequencing (SLAF-Seq). Front Plant Sci 12:688444. https://doi.org/10.3389/fpls.2021.688444

Liang Z, Zhang K, Chen K, Gao C (2014) Targeted mutagenesis in *Zea mays* using TALENs and the CRISPR/Cas system. JGG 41:63–68.

Liang K, Liang S, Lu L et al (2018a) Metabolic variation and cooking qualities of millet cultivars grown both organically and conventionally. Food Res Int 106:825–833. https://doi.org/10.1016/j.foodres.2018.01.023

Liang Z, Gupta SK, Yeh C-T et al (2018b) Phenotypic data from inbred parents can improve genomic prediction in pearl millet hybrids. G3 (Bethesda) 8:2513–2522. https://doi.org/10.1534/g3.118.200242

Liu M, Xu Y, He J et al (2016) Genetic diversity and population structure of broomcorn millet (*Panicum miliaceum* L.) cultivars and landraces in China based on microsatellite markers. Int J Mol Sci 17:370. https://doi.org/10.3390/ijms17030370

Lorenz K (1983) Tannins and phytate content in proso millets (*Panicum miliaceum*). Cereal Chem 60:424–426

Lu H, Zhang J, Liu K et al (2009) Earliest domestication of common millet (*Panicum miliaceum*) in East Asia extended to 10,000 years ago. Proc Natl Acad Sci U S A 106:7367–7372. https://doi.org/10.1073/pnas.0900158106

Majid A, Priyadarshini CGP (2020) Millet derived bioactive peptides: a review on their functional properties and health benefits. Crit Rev Food Sci Nutr 60:3342–3351. https://doi.org/10.1080/10408398.2019.1686342

Malik S (2015) Pearl millet-nutritional value and medicinal uses. IJARIIE 1(3):414–418

Matsushima N, Danno G, Takezawa H, Izumi Y (1997) Three-dimensional structure of maize a-zein proteins studied by small-angle X-ray scattering. Biochim Biophys Acta 1339:14–22

Nielsen DC, Calderón FJ (2011) Fallow effects on soil. In: Hatfield JL, Sauer TJ (eds) Soil management: building a stable base for agriculture. Soil Science Society of America, Madison, pp 287–300

Parameswaran KP, Thayumanavan B (1995) Homologies between prolamins of different minor millets. Plant Foods Hum Nutr 48:119–126. https://doi.org/10.1007/BF01088307

Rajput SG, Santra DK, Schnable J (2016) Mapping QTLs for morpho-agronomic traits in proso millet (*Panicum miliaceum* L.). Mol Breed 36:37. https://doi.org/10.1007/s11032-016-0460-4

Rao BD, Bhaskarachary K, Arlene Christina GD et al (2017) Nutritional and health benefits of millets, 1st edn. ICAR- Indian Institute of Millets Research, Hyderabad

Ravindran G (1992) Seed protein of millets: amino acid composition, proteinase inhibitors and in-vitro protein digestibility. Food Chem 44:13–17. https://doi.org/10.1016/0308-8146(92)90251-V

Rose DJ, Santra DK (2013) Proso millet (*Panicum miliaceum* L.) fermentation for fuel ethanol production. Ind Crop Prod 43:602–605. https://doi.org/10.1016/j.indcrop.2012.08.010

Roy SK, Kwon S-J, Yu J-H et al (2017) Comparison of protein profiles of proso millet (*Panicum miliaceum*) seeds of various Korean cultivars. Korean J Crop Sci 62:40–50. https://doi.org/10.7740/KJCS.2016.62.1.040

Saleh ASM, Zhang Q, Chen J, Shen Q (2013) Millet grains: nutritional quality, processing, and potential health benefits. Compr Rev Food Sci Food Saf 12:281–295. https://doi.org/10.1111/1541-4337.12012

Santra DK (2013) Proso millet varieties for Western Nebraska. University of Nebraska-Lincoln NebGuide, Scottsbluff, Nebraska

Santra DK, Rose DJ (2013) Alternative uses of proso millet. University of Nebraska-Lincoln Extension, p G2218

Santra DK, Heyduck RF, Baltensperger DD et al (2015) Registration of 'plateau' waxy (amylose-free) proso millet. J Plant Regist 9:41–43. https://doi.org/10.3198/jpr2013.11.0067crc

Santra DK, Khound R, Das S (2019) Proso millet (*Panicum miliaceum* L.) breeding: progress, challenges and opportunities. In: Al-Khayri JM, Jain SM, Johnson DV (eds) Advances in plant breeding strategies: cereals. Springer International Publishing, Cham, pp 223–257

Shevchenko A, Yang Y, Knaust A et al (2014) Proteomics identifies the composition and manufacturing recipe of the 2500-year old sourdough bread from Subeixi cemetery in China. J Proteome 105:363–371. https://doi.org/10.1016/j.jprot.2013.11.016

Shi J, Ma X, Zhang J et al (2019) Chromosome conformation capture resolved near complete genome assembly of broomcorn millet. Nat Commun 10:464. https://doi.org/10.1038/s41467-018-07876-6

Singh V, Moreau RA, Hicks KB (2003) Yield and phytosterol composition of oil extracted from grain sorghum and its wet-milled fractions. Cereal Chem 80:126–129. https://doi.org/10.1094/CCHEM.2003.80.2.126

Singh M, Adedeji A, Santra D (2018) Physico-chemical and functional properties of nine proso millet cultivars. Trans ASABE 61:1165–1174. https://doi.org/10.13031/trans.12553

Singode A, Balakrishna D, Bhat V et al (2018) Improving yield using EMS mutation in proso millet (*Panicum miliaceum* L). In: Santra DK, Johnson JJ (eds) The 3rd international broomcorn millet symposium. Fort Collins, Colorado, USA

Srivastava RK, Singh RB, Pujarula VL et al (2020) Genome-wide association studies and genomic selection in pearl millet: advances and prospects. Front Genet 10:1389. https://doi.org/10.3389/fgene.2019.01389

Tripathi MK, Mohapatra D, Jadam RS et al (2021) Nutritional composition of millets. In: Kumar A, Tripathi MK, Joshi D, Kumar V (eds) Millets and millet technology. Springer Singapore, Singapore, pp 101–119

Varshney RK, Shi C, Thudi M et al (2017) Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. Nat Biotechnol 35:969–976. https://doi.org/10.1038/nbt.3943

Vetriventhan M, Upadhyaya HD (2018) Diversity and trait-specific sources for productivity and nutritional traits in the global proso millet (*Panicum miliaceum* L.) germplasm collection. Crop J 6:451–463. https://doi.org/10.1016/j.cj.2018.04.002

Wang R, Wang H, Liu X et al (2018) Waxy allelic diversity in common millet (*Panicum miliaceum* L.) in China. Crop J 6:377–385.

Wietgrefe GW (1990) How to produce proso millet (a Farmer's guide). Wietgrefe, Gary W., Pierre, South Dakota

Yabe S, Iwata H (2020) Genomics-assisted breeding in minor and pseudo-cereals. Breed Sci 70:19–31. https://doi.org/10.1270/jsbbs.19100

Yue H, Wang L, Liu H et al (2016) De novo assembly and characterization of the transcriptome of broomcorn millet (*Panicum miliaceum* L.) for gene discovery and marker development. Front Plant Sci 7:1083. https://doi.org/10.3389/fpls.2016.01083

Zhang Y, Gao X, Li J et al (2019) Comparative analysis of proso millet (*Panicum miliaceum* L.) leaf transcriptomes for insight into drought tolerance mechanisms. BMC Plant Biol 19:397. https://doi.org/10.1186/s12870-019-2001-x

Zhao B, Khound R, Ghimire D et al (2022) Heading percentage estimation in proso millet (*Panicum miliaceum* L.) using aerial imagery and deep learning. Plant Phenome J. https://doi.org/10.1002/ppj2.20049

Zou C, Li L, Miki D et al (2019) The genome of broomcorn millet. Nat Commun 10:436. https://doi.org/10.1038/s41467-019-08409-5

Part II

Oilseed Crops

Nutraceuticals in Soybean: Biosynthesis, Advanced Genetic Research, and Usage in Food

Maria Stefanie Dwiyanti and Maria D. P. T. Gunawan-Puteri

Contents

M. S. Dwiyanti (✉)
Laboratory of Applied Plant Genome, Hokkaido University, Sapporo, Japan
e-mail: dwiyanti@agr.hokudai.ac.jp

M. D. P. T. Gunawan-Puteri
Department of Food Technology, Swiss German University, Tangerang, Indonesia
e-mail: maria.gunawanputeri@sgu.ac.id

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_12

Abstract

Soybean is a major protein feed and vegetable oil worldwide. Historically, soybean has been consumed in East and Southeast Asia, either as processed or fermented foods. Soybean also contains secondary metabolites such as isoflavones, saponins, and tocopherols (vitamin E) that are beneficial for human health. Studies have been conducted to understand the genetic basis of their biosynthesis and accumulation. Content and composition of most compounds change as response to environmental conditions during the plant growth and seed filling. This chapter will describe findings on oil, protein, isoflavones, saponins, and tocopherol biosynthesis and accumulation in soybean seeds. The availability of next-generation sequencing technologies has changed the way genetic analysis is done, and have accelerated the elucidation of genetic basis of the nutritional components. Moreover, the NGS affordability has provided us with reference genomes, large-scale sequencing data, and transcriptome data that are publicly available. Researchers can access and utilize this data to complement their findings. Lastly, this chapter will describe soy-based foods, the processing and nutritional values, and their potential as nutritional source and health component resources.

Keywords

Fatty acid · Functional food · Isoflavones · Protein · Saponin · Soybean · Tocopherols · Vitamin E · Wild soybean

1 Introduction

Seeds of soybean (*Glycine max* (L.) Merrill) contain oil (20% of seed weight), protein (40% of seed weight), and secondary metabolites (isoflavones, saponins, and tocopherols). Having high protein and oil content, soybean became one of the major sources of protein and oil. During 2020/2021, soybean protein meal consumption as feed was 243.6 million metric tons, which contributed to 71% of the total protein meal consumed worldwide (Soystats 2022). Soybean oils contributed to 29% (58.7 million metric tons) of world oil consumption (Soystats 2022). Traditionally, soybean has been consumed in East Asia and Southeast Asia as various foods, which can be categorized into two, fermented and nonfermented foods. Nonfermented foods examples are tofu, bean sprout, soymilk, and yuba. Examples of fermented foods are natto, tempeh, miso, jang, and soy sauce. The fermentation process changes the nutritional components of soybean foods, reducing antinutrient components and increasing nutritional compounds. Natto and tempeh contain high vitamin K2 and B12, which are not detected in raw soybean seeds. The fermentation process also adds flavor through the catabolism of protein to amino acids and

reduces antinutritional components such as lipoxygenases, trypsin inhibitors, and phytic acid (Nout and Kiers 2005). In addition to traditional food, recent interest in plant-based diets also increased the attention to soy-based food. Furthermore, the potential of soybean as nutritious food for children, a new protein source in Africa and Europe, and the use of by-products as a new source of nutrients is being explored.

The domestication of soybean occurred about 6000–9000 years ago in Yellow River, Central China (Kofsky et al. 2018). The wild ancestor of soybean is *G. soja* (Siebold &Zucc.). Wild soybean naturally grows in diverse habitats in East Russia, Japan, China, and Korea (Kofsky et al. 2018). Wild soybean plants form vines, giving it the Japanese name "tsurumame," which means "bean (plants) with vines." The seeds are small and black. Soybean and wild soybean outcrossing produce fertile progenies. According to Kofsky et al. (2018), wild soybean has larger genetic diversity compared to soybean. Moreover, wild soybean possesses unique genes not available in soybean (Kofsky et al. 2018). Thus, wild soybean provides a huge genetic reservoir for current soybean cultivars' improvement. To tap this potential, screening natural variants of metabolites content and resequencing projects also include wild soybean in the analysis.

2 Nutritional Components: Biosynthesis and Regulation

2.1 Oil and Fatty Acids

Soybean oils contribute to 29% (58.7 million metric tons) of world oil consumption (Soystats 2022). Oil content per seed dry weight of soybean seeds is approximately 20%. It varies depending on varieties, as well as growth conditions (Clemente and Cahoon 2009). Soybean oil is composed by mainly five fatty acids, palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), and linolenic acid (C18:3) (Clemente and Cahoon 2009). Generally, linoleic acid is the most abundant fatty acid in soybean oil, about 55% of the total fatty acid content (Clemente and Cahoon 2009). Among five fatty acids, oleic acid, linoleic acid, and linolenic acid are categorized as unsaturated fatty acids, because they have double bonds in their carbon chains. The other two, palmitic acid and stearic acid, are saturated fatty acids, having no double bonds in their carbon chains. Carbon chain length and the level of unsaturation are shown as numbers after "C." For example, oleic acid is described as C18:1, meaning it has 18 carbons and one double bond. Unsaturated fatty acids are more easily oxidized compared to saturated fatty acids, and if the fatty acids have higher number of double bonds, they are less stable. Therefore, linolenic acid is the most easily oxidized, followed by linoleic acid, and oleic acid. Oxidation makes oil becomes rancid and reduces its shelf life. Partial hydrogenation reduces the proportion of unsaturated fatty acids and creates trans-fatty acids, which are linked to cardiovascular diseases (Clemente and Cahoon 2009).

Linoleic (C18:2) and linolenic acids (C18:3) are essential dietary elements for humans and higher animals. They are also able to decrease the risk of coronary heart

diseases (Clemente and Cahoon 2009). Since human body cannot produce linoleic acid and linolenic acid, breeding cultivars with the optimum composition of unsaturated fatty acid and saturated fatty acid is one of the soybean breeding objectives. Quantitative trait locus (QTL) and genome-wide association study (GWAS) analyses have identified numerous quantitative trait loci (QTLs) associated with oil and fatty acid content in soybean seeds (Patil et al. 2017, 2018). Some of the oil content QTLs shared position with protein content but with contrasting effect.

2.1.1 Fatty Acids Biosynthesis and Its Regulation

Fatty acids are synthesized from acetyl-CoA in the plastid. Acetyl-CoA is carboxylated to produce malonyl-CoA. Fatty acid transferase adds two carbon atoms to the acyl chain on acyl carrier protein (ACP) repeatedly to create 16 carbon acyl-ACP (C16:0-ACP). Reduction, dehydration, and reduction processes also occur along with the elongation process. C16:0-ACP can be elongated further by ketoacyl-ACP synthase II (KASII) to produce C18:0-ACP. Stearoyl-ACP desaturase desaturases C18:0-ACP to produce C18:1-ACP. Free fatty acids (C16:0, C18:0, and C18:1) are released from ACP and transported to the endoplasmic reticulum (ER). In ER, C18:1 is further desaturated by fatty acid desaturases to C18:2, and lastly to C18:3. Fatty acids in soybean seeds are stored in the form of triacylglycerol (TAG). Free fatty acids from ER and plastid are incorporated into glycerol-3-phosphate and undergo a series of conversions to create diacylglycerol (DAG). DAG is converted to triacylglycerol (TAG) by acyl-CoA:DAG acyltransferase (DGAT). TAG in soybean seeds is mostly stored in oil bodies (oleosomes), within membranes containing phospholipids and oleosins.

Fatty acid desaturase-2 (FAD2) desaturases oleic acid (C18:1) to linoleic acid (C18:2). There are five *FAD2* genes in soybean. Mutations at either *GmFAD2-1A* or *GmFAD2-1B* increased oleic acid content to 80% and decreased linoleic acid content to 30% (Anai et al. 2008). Double mutations of *GmFAD2-1A* or *GmFAD2-1B* produced lines having 80% oleic acid whereas linoleic acid was reduced to less than 5% (Pham et al. 2012). Desaturation of linoleic acid (C18:2) to linolenic acid (C18:3) is catalyzed by FAD3. There are three genes encoding FAD3, namely *GmFAD3A*, *GmFAD3B,* and *GmFAD3C* (Pham et al. 2012). Combination of *GmFAD3A-GmFAD3B* or *GmFAD3A-GmFAD3C* mutations produced lines having linolenic acid about 3% of total oil content, whereas triple mutations of *GmFAD3* genes resulted in lines having 1% linolenic acid content (Pham et al. 2012). Pham et al. (2012) further combined *GmFAD2-1* double mutants with one *GmFAD3* mutant, producing mutants having less than 2% linolenic acid. Interestingly, growing locations affected the linolenic acid content of the mutants (Pham et al. 2012). Lines having triple mutants (*FAD2-1aabb FAD3aaCC* or *FAD2-1aabb FAD3AAcc*) had linolenic acid content less than 3% when they were grown in Portageville, Missouri, but quadruple mutants (*FAD2-1aabb FAD3aacc*) type is needed to achieve the same level of linolenic acid content for growing in Columbia (Missouri) which is located north of Portageville (Pham et al. 2012). Demorest et al. (2016) used transcription activator–like effector nucleases (TALEN), a gene editing method, to mutate

GmFAD3A in *fad2-1a- fad2-1b* mutants and produced lines with low linolenic acid (2.5%) and high oleic acid (82.2%).

2.2 Protein

Protein contributes to about 40% of seed dry weight in commercial cultivars, and varies based on genotype or growth environment. Multiple genes regulate protein content. More than 160 QTLs distributed across 20 chromosomes have been identified from 35 independent studies (Patil et al. 2017). Of these, QTLs located on chromosome 15 and 20 are strongly correlated with protein content. The result was also confirmed in large-scale GWAS analysis conducted on 12,000 soybean accessions (Bandillo et al. 2015). Candidate gene analysis on chromosome 15 and 20 both identified three candidate genes each, which need further investigation for their roles in protein content regulation (Bandillo et al. 2015). Both QTLs also regulated oil content (Bandillo et al. 2015), but total protein content is negatively correlated with oil content (Patil et al. 2017). Therefore, it is challenging to breed a soybean cultivar having both high protein and oil content. While genetic breeding remains focused on seed yield and soybean oil remains as the major consumption of global soybean production, there might be a very few cultivars developed for its ultrahigh protein content though soybean accessions with >50% protein content were reported (Patil et al. 2017).

Protein composition is also important for soybean nutritional value. About 70% of total storage proteins in soybean is composed by glycinin (11S globulin) and β-conglycinin (7S globulin) (Takahashi et al. 2003). β-conglycinin (7S globulin) is a heterotrimeric protein, and has three subunits: α, α′, and β-subunit. β-subunit has 420 amino acids (AA) that are shared with α and α′-subunit. The α and α′-subunits have additional 125 AA and 141 AA in the N-terminal, respectively. Glycinin (11S globulin) is a heterohexameric protein, consisting of five subunits A1aB1b, A1bB2, A2B1a, A3B4, and A5A4B3 (Adachi et al. 2003). Between two storage proteins, the amount of methionine and cysteine per unit protein is higher in 11S globulin (Kitamura 1995). Methionine is an essential amino acid that cannot be synthesized in human body but it is important to build proteins and molecules in human body. Sufficient intake of methionine is important for the proper function of cells. Moreover, high 11S to 7S globulin is preferred in tofu processing, as it forms harder curd in tofu production (Kitamura 1995).

Several breeding programs have aimed to increase the ratio of 11S globulin to 7S globulin using natural mutants in germplasm or mutants gained by γ-ray irradiation (Kitamura 1995). For example, 'Kebuli' lacking α'-subunit and 'Moshidou Gong 503' having low levels of α- and β-subunits have been used to breed lines with low α, β-subunits, and null α'-subunit (Kitamura 1995). The 7S-low lines had 50% lower 7S globulin and the 11S globulin was 15% higher compared to normal varieties without change in total protein content (Kitamura 1995). Gamma-ray irrradiation of 'Karikei 434' produced a mutant line lacking both α- and β-subunits and low level of α'-subunit without decrease in total protein content or defect in plant development

(Takahashi et al. 1994). The reduced 11S globulin content in soybean were compensated by increase in 7S globulin (Yang et al. 2016). Takahashi et al. (2003) produced a mutant line lacking both 7S and 11S globulin without any defect in growth and production by crossing a line lacking 7S globulin with a line lacking 11S globulin. Protein bodies in cotyledons of the mutant is underdeveloped but the nitrogen content did not change compared to wild-type cultivars (Takahashi et al. 2003). Interestingly, free amino acids (arginine, asparagine, glutamic acid, and histidine) increased in the mutant, contributing to 4.5–8% of seed nitrogen content, whereas the proportion in wild-type varieties is between 0.3% and 0.8% (Takahashi et al. 2003).

Soybean also has proteins which are allergens (Wilson et al. 2005). Subunits of 7S globulin also generate antibody response in mice fed with soy protein (Wilson et al. 2005). There are about 20 proteins identified as allergen in soybean. Three known major soybean allergens are *Gly m* Bd 60 K (α-subunit of β-conglycinin), *Gly m* Bd 30 K, and *Gly m* Bd 28 K (vicilin-like glycoprotein), respectively. P34 (Gly m Bd 30 K) shares 70% homology with peanut main allergen (Wilson et al. 2005). Due to this, most patients having allergy to peanuts also showed allergy to soybean (Wilson et al. 2005). Fermentation process hydrolyzes proteins and as the result, fermented soybean products such as miso, natto, tempeh, and soy sauce are potentially less allergenic than raw soybeans. For example, *Gly m* Bd 28 K content is reduced in fermented soybean products compared to raw soybean and nonfermented soy products (Ogawa et al. 2000; Bando et al. 1998). Another example of reducing soybean allergens is through breeding, such as low-P34 varieties having reduced P34 (Gly m Bd 30 K) content (Bilyeu et al. 2009). Analysis of soybean germplasm identified natural mutants such as PI 567476 and PI 603570A, which contain four base pairs insertion at P34 start codon resulting in translation initiation frameshift of the protein (Bilyeu et al. 2009). Molecular markers were developed to recognize this mutation and were used in marker-assisted selection for low P34 soybean cultivars (Watanabe et al. 2017).

2.3 Isoflavones

Isoflavones are metabolites derived from phenylalanine pathway. They are mainly found in legumes (Dhaubhadel et al. 2003, 2007). Soybean isoflavones can be categorized to three types based on their aglycone structure, daidzein, genistein, and glycitein. The aglycone forms are biologically active and are absorbed in the human intestine. Isoflavones have many human benefits, including reducing hormone-dependent cancers and cardiovascular disease risk alleviating postmenopausal symptoms, and preventing osteoporosis. In plants, it induces phytoalexin production as a response to pathogen attack, induces the expression of nodulation genes, and regulates the formation of nodules in soybean roots (Dhaubhadel et al. 2003).

Two major of soybean isoflavones are exist as malonylglycoside and glycoside forms (Tsukamoto et al. 1995). Isoflavones are synthesized starting at an early development stage and the accumulation rapidly increases at later stages of seed

development (Dhaubhadel et al. 2003, 2007). Seed isoflavone content is affected by the environment, particularly the temperature during plant growth and seed maturation (Chennupati et al. 2011). Seven soybean varieties were grown in Kyushu and six varieties were grown in Tsukuba in three sowing times: April, May, and June 1991 (Tsukamoto et al. 1995). Average temperatures in Tsukuba and Kyushu differed by 4 °C in July–August 1991, but the average temperatures differed only 1 °C in October 1991 (Tsukamoto et al. 1995). Four varieties ('Koganedaizu', 'Shirosaya I', 'Higomusume', and 'Kairyoshirome') belonged to the early maturity group (Group I and II) and were known as low isoflavone varieties, and the other three varieties grown as control belonged to maturity group IV and VI (Tsukamoto et al. 1995). Interestingly, regardless of the varieties, malonyldaidzin and malonylgenistin contents were higher in whole seeds of the plants sown in July 1991 compared to seeds of plants from earlier sowing dates (Kyushu only). Harvest dates of the early maturity group sown in July 1991 were late September 1991, when the average temperature was 24.7 °C, the maximum was 29.5 °C, and the minimum was 19.8 °C (Tsukamoto et al. 1995), which was lower compared to an average of 27.2 °C in August 1991 (Tsukamoto et al. 1995). Chennupati et al. (2011) also investigated the effect of high temperature on the isoflavone content of two cultivars, 'AC Proteina' (high isoflavone) and 'OAC Champion' (low isoflavone). Stress condition of 33 °C/25 °C (day/night temperature) and control condition of 23 °C/15 °C (day/night temperature) were imposed during all developmental stages, pre-emergence, vegetative stage, early seed filling (R1-R4 stages), and late seed filling stage (R5-R8 stages) (Chennupati et al. 2011). Both cultivars showed a reduction in total isoflavone content when stress was imposed during all development stages and late seed filling stages (Chennupati et al. 2011), which coincided with the rapid isoflavone accumulation during late seed filling stages.

2.3.1 Isoflavone Biosynthesis and Its Regulation

Isoflavone biosynthesis is a part of the phenylpropanoid pathway, which is also a precursor pathway for anthocyanins and lignin biosynthesis (Fig. 1). Phenylalanine ammonia lyase (PAL) catalyzes deamination of phenylalanine to cinnamic acid as the first enzyme of isoflavone biosynthesis process. Cinnamic acid is converted to p-coumaryol CoA through two steps catalyzed by cinnamate 4-hydroxylase (C4H) and 4-coumarate CoA ligase (4CL). P-coumaryol CoA and its precursor, p-coumaric acid, are a substrate for lignin biosynthesis. The next step is catalyzed by chalcone synthase (CHS). CHS converts p-coumaryol CoA to naringenin chalcone, and together with chalcone reductase converts p-coumaryol CoA to isoliquiritigenin. Naringenin chalcone and isoliquiritigenin are further converted to naringenin and liquiritigenin, respectively. This step is catalyzed by chalcone isomerase. Naringenin is converted to genistein by isoflavone synthase (IFS). IFS also catalyzes the conversion of liquiritigenin to daidzein. Glycitein is produced from liquiritigenin by a series of conversions involving flavonoid 6-hydroxylase (F6H) and IFS. Genistein, daidzein, and glycitein are further converted to their glucosides from (genistin, daidzin, and glycitin) by isoflavone 7-*O*-glucosyltransferase (IF7GT). Lastly, the malonylglucosides are synthesized from isoflavone glucosides, and the reaction is catalyzed by isoflavone 7-O-glucoside 6"-*O*-malonyltransferase

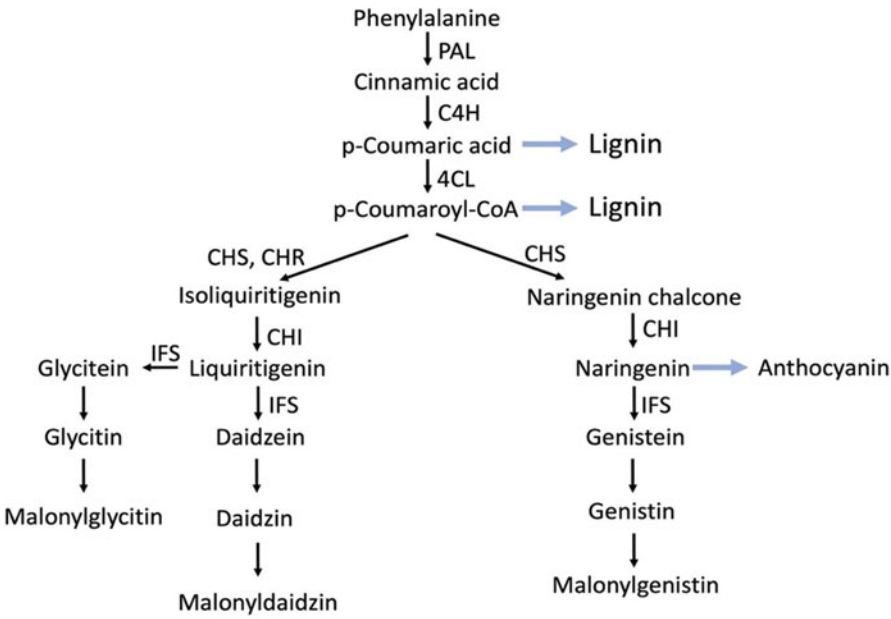


Fig. 1 Isoflavones biosynthesis pathway. Abbreviations for enzymes: *PAL* phenylalanine ammonia lyase, *C4H* cinnamate 4-hydroxylase, *C4L* 4-coumarate CoA ligase, *CHS* chalcone synthase, *CHR* chalcone reductase, *CHI* chalcone isomerase, and *IFS* isoflavone synthase

(IF7MaT) (Fig. 1). CHS and IFS are the most studied among enzymes in the isoflavone biosynthesis pathway.

Soybean contains multiple genes encoding for CHS. Dhaubhadel et al. (2007) performed transcriptome array analysis on two soybean cultivars, 'Harovinton' (low isoflavone) and 'RCAT Angora' (high isoflavone), to determine genes involved in isoflavone biosynthesis. The transcriptome analysis was conducted on five stages of seed development (Dhaubhadel et al. 2007). Compared to other copies, *CHS7* and *CHS8* are expressed in seeds and the expression level increased toward seed maturity (Dhaubhadel et al. 2007). The expression increases also coincided with the accumulation of seed isoflavones (Dhaubhadel et al. 2007).

There are two gene copies for IFS: *IFS1* and *IFS2*. Both copies are involved in isoflavone biosynthesis, but their transcription is regulated differently (Dhaubhadel et al. 2003). The genes are very similar, only *IFS1* is expressed in all tissues, but its expression is the highest in root and seed coats (Dhaubhadel et al. 2003). *IFS2* expression is low in stems, leaves, pods, and seed coats (Dhaubhadel et al. 2003). Its expression is high in embryos, developing seeds, and late-stage pods (Dhaubhadel et al. 2003). Interestingly, the expression level of *IFS2* in developing seeds increases toward maturity, whereas the *IFS1* expression is constant (Dhaubhadel et al. 2003). On the other hand, *IFS2* plays role in response to pathogen attack since its expression increased in hypocotyls and roots after pathogen attacks (Dhaubhadel et al. 2003; Subramanian et al. 2004).

Transcription levels of isoflavone biosynthesis genes are regulated by numerous transcription factors (TFs) such as MYB, bHLH, bZIP, WRKY, and MADS-boxand WD40 (Yi et al. 2010). There are more than 5000 genes putatively encoding TFs, and MYB TFs comprise about 14% of TFs. Based on the number of MYB domains, MYBs are classified into 1R, 2R (R2R3-type), 3R, and 4R MYBs. *GmMYB176* encoding R1-type MYB was identified as a regulator for *CHS8* (Yi et al. 2010). Cotransfection assay showed that *GmMYB176* transactivated the *CHS8* promoter (Yi et al. 2010). Hairy roots RNAi-mediated gene silencing of *GmMYB176* reduced isoflavone content, but the overexpression did not increase the transcript level of CHS8 (Yi et al. 2010). R2R3-type MYB TFs such as *GmMYB100*, *GmMYB39,* and *GmMYB29* were also identified as potential regulators for isoflavone biosynthesis (reviewed in Sohn et al. 2021).

Although genes encoding biosynthesis genes and transcription factors regulating isoflavone content have been identified, the regulation in natural germplasm is much more complex, since genotype × environment interaction effect is large, and flavonoids are also involved in biotic stress response. More than 200 QTLs distributed in 20 chromosomes have been reported from a GWAS and QTL analyses using 200 soybean cultivars and 150 RILs identified several SNPs and QTLs associated with isoflavone content (Wu et al. 2020).

2.4 Saponins

Soyasaponins are triterpenoid glycosides (Sundaramoorthy et al. 2018). In soybeans, there are two major types of saponins, group A saponins and DDMP saponins (Sundaramoorthy et al. 2018). Group A saponins are the major form of saponin in hypocotyl, whereas DDMP saponins is a predominant form in cotyledon (Sundaramoorthy et al. 2018). Group A saponins cause bitter and astringent taste in soy food products (Tsukamoto et al. 1995). On the other hand, DDMP saponin showed health benefits such as inhibition of HIV infection and activation of Epstein-Barr virus early antigen (Tsukamoto et al. 1995). Therefore, decreasing the amount of group A saponins and increasing DDMP saponins would contribute to the improvement of soy-based foods' quality (Tsukamoto et al. 1995).

Group A saponins and DDMP saponins differ in the C-21 position of their soyasapogenol structure. Group A saponins have a hydroxy group (-OH) at the C-21 position (Fig. 2). The aglycones are called soyasapogenol A orsoyasapogenol B (Fig. 3). DDMP saponins have a 2,3-dihydro-2,5-dihydroxy-6-methyl-4H-pyran-4-one (DDMP) moiety attached to the C-22 position of the soyasapogenol B (Fig. 3). During food processing, DDMP saponins are degraded to B saponin and E saponin (Fig. 3, Krishnamurthy and Min 2014). In addition, there are one or two sugar moieties attached to the C-3 position (R_1), and acetylxylose or acetylglucose attached to the C-22 position (R_2) resulting in saponin variations (Fig. 2). Saponins are derived from β-amyrin, which is a product of the mevalonate pathway (Sundaramoorthy et al. 2018). Genes involved in soybean saponin biosynthesis

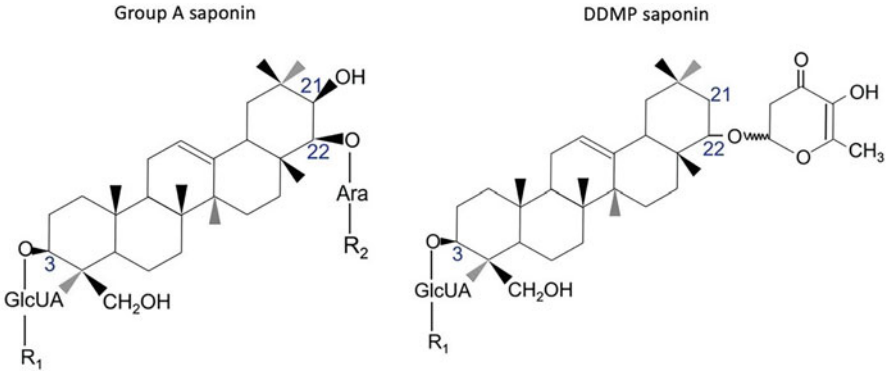


Fig. 2 Group A saponin and DDMP saponin structure. R_1 is either galactose (Gal), arabinose (Ara), combinations of glucose-galactose (Glc-Gal), rhamnose-galactose (Rha-Gal), Glc-Ara, or Rha-Ara. R_2 is either acetylglucose or acetylxylose

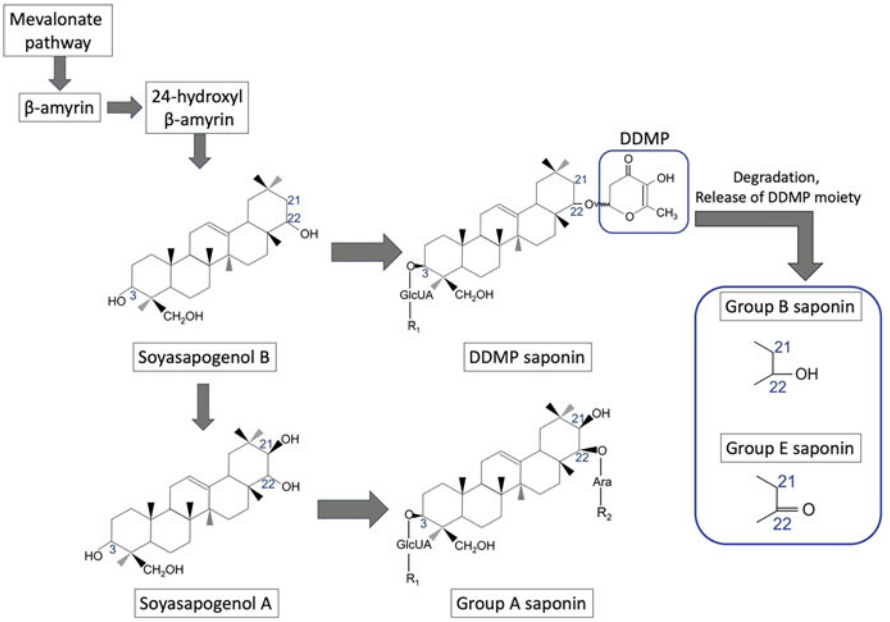


Fig. 3 Saponin biosynthesis pathway

pathway has not yet been fully elucidated, but several mutants having altered saponin composition have been identified (Fig. 3).

Two mutants, *Sg-1*^{a}and *Sg-1*^{b}, showed a mutation in Glyma.07G254600, encoding a glycosyltransferase (Sayama et al. 2012). The significant difference between them is the amino acid number 138; Sg-1^a and Sg-1^b proteins have serine and glycine, respectively (Sayama et al. 2012). This resulted in a difference in their

function. Sg-1^{a} adds xylose whereas Sg-1^{b}adds glucose to the aglycone (Sayama et al. 2012). Accessions with the *Sg-1^{a}*allele have acetylxylose, and accessions with the *Sg-1^{b}* allele have acetylglucose as the third sugar group at the C-22 position. Mutants having loss-of-function allele *sg-1^{0}* accumulated saponin A0-αg lacking acetylated terminal sugar at the C-22 position (Sayama et al. 2012). The *sg-1^{0}* mutant has been used to breed 'Kinusayaka', which has a less bitter taste and astringent flavor. The variety is suitable for soymilk and tofu production (Kato et al. 2007).

Sg-3 locus contains a glucosyltransferase UGT91H4 (Glyma.10G104700). The enzyme adds glucose as a third sugar at the C-3 position of group A saponin or DDMP saponin. *Sg-4* locus contains a glycosyltransferase UGT73P10 (Glyma.01-G046300). Homozygous recessive *sg-4* mutants lacked saponins that have arabinose as the second sugar group at the C-3 position (Takagi et al. 2018).

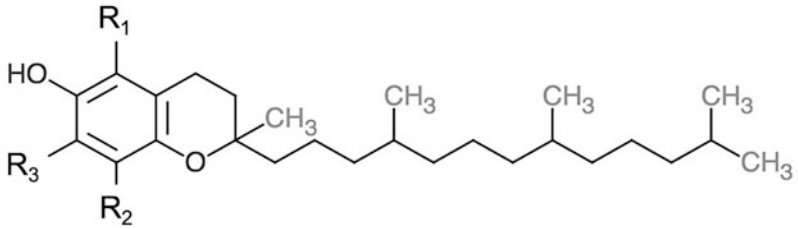
Sg-5 (Glyma.15g243300) encodes a cytochrome P450 enzyme (Yano et al. 2017). The natural *sg-5* mutant was first identified in wild soybean, and the mutant did not accumulate group A saponin but showed a high level of DDMP saponin (Yano et al. 2017). This mutant had a premature stop codon in Glyma.15g243300 (Yano et al. 2017). Since high DDMP-low group A saponin is desirable for food production, the *sg-5* allele has been introduced to 'Tohoku 152' (Yano et al. 2017). 'Tohoku 152' has low amount of group A saponins both in hypocotyl and cotyledon. Compared to cultivars carrying *Sg-5*, 'Tohoku 152' has higher DDMP saponin in the hypocotyl (Yano et al. 2017).

2.5 Tocopherols (Vitamin E)

Tocopherols are known as vitamin E. Tocopherol is consisted of a chromanol head and a saturated phytyl side chain. The number and position of the methyl groups on the chromanol head determine the isoforms as α-, β-, γ-, and δ- tocopherol (Fig. 4). Among four isoforms, α-tocopherol has high affinity with tocopherol transfer protein in the human liver, and kept at a high level in blood plasma. Thus, the vitamin E activity of α-tocopherol is the highest among four isoforms. The vitamin E activities of β-, γ-, and δ-tocopherol are 0.5, 0.1, and 0.03 when compared to vitamin E activity of α-tocopherol (equal to 1), respectively (Van Eenennaam et al. 2003).

Total tocopherol content of soybean seed oil is higher than other oils from canola, sunflower, or palm. The major isoform in seeds is γ-tocopherol (60–70%), followed by δ-tocopherol (20–30%), whereas α-tocopherol content is less than 10%. Increasing α-tocopherol content in soybean seeds may improve the vitamin E status of soybean. On the other hand, γ-tocopherol and its derivative showed an anti-inflammatory effect, which is a specific function that was not observed in α-tocopherol (Jiang et al. 2001). Altering the ratio of γ-tocopherol and α-tocopherol may be preferred instead of increasing only the α-tocopherol content.

Seed tocopherol content diversity of soybean germplasm has been surveyed in Japan (Ujiie et al. 2005; Dwiyanti et al. 2016), India (Rani et al. 2007), and Brazil (Carrão-Panizzi and Erhan 2007). From the screening of more than 1000 soybean and wild soybean accessions, three accessions containing high α-tocopherol



Compound	R1	R2	R3	Vitamin E Activity (ratio to α-tocopherol)
α-Tocopherol	CH_3	CH_3	CH_3	1
β-Tocopherol	CH_3	H	CH_3	0.5
γ-Tocopherol	H	CH_3	CH_3	0.1
δ-Tocopherol	H	H	CH_3	0.03

Fig. 4 Tocopherol structure and isoforms

(more than 20% of total tocopherol content) were identified: 'Keszthelyi Aproszemu Sarga', 'Dobrogeance', and 'Dobrudza 14 Pancevo' (Ujiie et al. 2005). The α-tocopherol content ranged from 58 μg/g oil to 794 μg/g oil (Rani et al. 2007). The highest α-tocopherol percentage (27%) was observed in the variety 'Ankur' (Rani et al. 2007). The α-tocopherol content 89 Brazil soybean accessions varied between 11 ppm ('Davis') and 191 ppm ('IPB-T') (Carrão-Panizzi and Erhan 2007). 'IPB-T' also showed high total tocopherol content (1386.29 ppm) (Carrão-Panizzi and Erhan 2007).

Tocopherols accumulate during the later stage of soybean seed development (R5–R8 stages). High temperature (28 °C) and mild drought during these stages increased seed α-tocopherol content approximately twofold compared to those grown at 23 °C (Britz and Kremer 2002). When the high temperature was imposed in different developmental stages, high temperature during whole growth stages or seed filling (R5–R8 stage) gave the largest effect on the increase in α-tocopherol content (Chennupati et al. 2011). In high temperatures, δ-tocopherol content decreased (Britz and Kremer 2002; Chennupati et al. 2011). Total tocopherol content and γ-tocopherol content change varied depending on studies and varieties (Britz and Kremer 2002; Chennupati et al. 2011).

2.5.1 Tocopherol Biosynthesis and Its Regulation

The tocopherol biosynthesis pathway in higher plants is already elucidated (Fig. 5). The first component in the tocopherol biosynthesis pathway is 2-methyl-6-phytyl-1,4-benzoquinone (MPBQ) which is synthesized by combining homogentisic acid (HGA) and phytyl-diphosphate (PDP). HGA is produced from the shikimate pathway, whereas PDP is produced from the MEP pathway or chlorophyll degradation

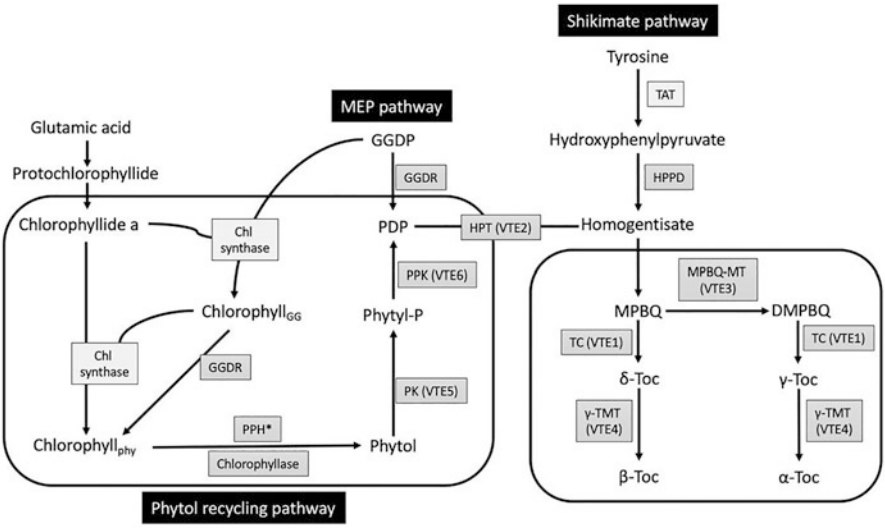


Fig. 5 Tocopherol biosynthesis and chlorophyll degradation/phytol recycling pathway. MEP: methylerythritol phosphate. Abbreviations for substrates: *GGDP,* geranylgeranyl diphosphate; *PDP,* phytyl diphosphate; *MPBQ,* methylphytylbenzoquinol; *DMPBQ,* dimethylphytylbenzoquinol; *Toc,* tocopherol. Abbreviations for enzymes: *GGDR,* geranylgeranyl diphosphate reductase; *PPK,* phytyl phosphate kinase; *PK,* phytyl kinase; *Chl synthase,* chlorophyll synthase; *PPH,* pheophytin pheophorbide hydrolase; *TAT,* tyrosine aminotransferase; *HPPD,* hydroxyphenylpyruvate dioxygenase; *HPT,* homogentisatephytyl transferase; *MPBQ-MT,* MPBQ methyltransferase; *TC,* tocopherol cyclase; *γ-TMT,* γ-tocopherol methyltransferase

pathway. MPBQ is subsequently methylated by MPBQ- methyltransferase enzyme (MPBQ-MT) and is converted to 2,3-dimethyl-5- phytyl-1,4-benzoquinone (DMPBQ). The δ-tocopherol and γ-tocopherol are produced from cyclization of the chromanol heads of MPBQ and DMPBQ, respectively. Tocopherol cyclase is involved in this process. Gamma-tocopherol methyltransferase (γ-TMT) adds a methyl group to the chromanol head of δ-tocopherol and γ-tocopherol, converting them to β-tocopherol and α-tocopherol, respectively. Most tocopherol biosynthesis genes were first identified in Arabidopsis. The names of the corresponding Arabidopsis genes (*VTE* genes) are annotated in Fig. 5. MPBQ-MT and γ-TMT enzymes play important role in determining tocopherol composition in soybean seeds (Van Eenennaam et al. 2003). Overexpression of Arabidopsis *VTE4* gene expressing γ-TMT (Van Eenennaam et al. 2003) could increase seed α-tocopherol percentage up to 50–75% of total tocopherol content, and reduced δ-tocopherol and γ-tocopherol percentage. Overexpression of Arabidopsis *VTE3* gene encoding the MPBQ-MT reduced δ-tocopherol and β-tocopherol, and increased γ-tocopherol and α-tocopherol 10% compared to wild type. Coexpression of both *VTE3* and *VTE4* expressing MPBQ-MT and γ-TMT further improved α-tocopherol up to 60.4–91%

of total tocopherol content, eliminated most of δ-tocopherol and β-tocopherol, and reduced γ-tocopherol (Van Eenennaam et al. 2003).

Overexpression of the Arabidopsis *VTE2* gene expressing homogentisate phytyltransferase (HPT) resulted in slight increase of total tocopherol content (Karunanandaa et al. 2005). Seed-specific overexpression of four genes encoding Arabidopsis HPPD, VTE2, GGDP dehydrogenase (GGH), and *Erwinia herbicola* prephenate dehydrogenase (*Eh-TYRA*) was conducted (Karunanandaa et al. 2005). TYRA catalyzes the synthesis of p-hydroxyphenylpyruvate, homogentisic acid precursor. The overexpression of four genes increased tocochromanols up to 15-fold in the best transgenic event (Karunanandaa et al. 2005). Surprisingly, up to 94% of tocochromanols in transgenic seeds were tocotrienols instead of tocopherols (Karunanandaa et al. 2005). How the combination of four genes led to conversion to tocotrienols is still unknown.

Based on the Williams82 gene annotation, soybean has multiple gene copies of each enzyme in tocopherol biosynthesis. The copies may show differentiation in function and response to the growth environment. For example, there are three gene copies for γ-TMT: Glyma.12G014200 (*γ-TMT1*), Glyma.12G014300 (*γ-TMT2*), and Glyma.09G222800 (*γ-TMT3*) (Dwiyanti et al. 2011). Based on public transcriptomics data, the three copies are expressed in leaves, flowers, and early-stage pods (Severin et al. 2010; Le et al. 2007). Since tocopherol also plays role as antioxidant in photosynthetic organs, the gene expression observed in leaves, flowers, and pods may relate to its function. Interestingly, in developing seeds, the difference in expression pattern was observed. The *γ-TMT1* is expressed higher in early stage of developing seeds and gradually decreased toward maturation. On the other hand, the seed *γ-TMT2* and *γ-TMT3* expression level increases toward seed maturation. Among three copies, only *γ-TMT2* expression level is elevated at high temperature, whereas *γ-TMT1* and *γ-TMT3* did not increase (Park et al. 2019).

Glyma.09G222800 (*γ-TMT3*) was identified as a candidate gene responsible for high α-tocopherol content in KAS. The candidate gene was identified based on QTL analysis on an F_5 population derived from 'Ichihime' × 'KAS' cross (Dwiyanti et al. 2011). Further confirmation using GUS-reporter assay in leaves of transgenic Arabidopsis transformed with 'KAS' *γ-TMT3* promoter-intron GUS and with 'Ichihime' *γ-TMT3* promoter-intron GUS showed that 'KAS' *γ-TMT3* promoter activity was higher than that of 'Ichihime' (Dwiyanti et al. 2011). Another QTL analysis using recombinant inbred lines (RILs) population derived from a Hokkaido cultivar 'TK780' and high α-Toc wild soybean ('B04009') identified several QTLs, including a QTL containing *γ-TMT3* and a QTL containing *γ-TMT1* and *γ-TMT2* (Park et al. 2019). Interestingly, both studies found *γ-TMT3* as the candidate gene and same single-nucleotide polymorphisms (SNPs) differentiating between high and low varieties (Dwiyanti et al. 2011; Park et al. 2019). One of these was located within CAAT-box, which is located 74-bp upstream the translation start codon (Dwiyanti et al. 2011). Compared to 'TK780' and 'Ichihime', the *γ-TMT3* expression level was elevated in 'KAS' and 'B04009', during seed maturation. This correlated to difference in α-tocopherol content (Dwiyanti et al. 2011; Park et al. 2019).

3 Genetic Marker Resources and Genotyping Technologies

3.1 RFLP, AFLP, and RAPD

Restriction fragment length polymorphism (RFLP) is a marker based on DNA variations within restriction enzyme sites. Restriction enzymes (REs) recognize RE sites and cut the DNA. If DNA variation occurs at the RE sites, RE cannot cut the DNA. RE then can be used to digest DNA, and the differences in fragment length are the source of RFLP markers. RFLP is a codominant marker and does not need reference genome information, therefore it is suitable for soybean genotyping before the whole genome sequence was available. The first soybean genetic linkage map consisted of 26 linkage groups and was constructed from 150 RFLP markers based on a F_2 population derived from a cross between a soybean cultivar A81-356022 and a wild soybean accession PI468916 (Song et al. 2004). After that several RFLP-based linkage maps were constructed (Song et al. 2004). However, RFLP has disadvantages: large amount of DNA needed for digestion and radioactive probe usage to detect polymorphism. In soybean, the polymorphism rate of RFLP markers is low. Also, since soybean is an ancient polyploid, RFLP probes map to more than one position in the genome, makes it difficult to compare the genotyping results across different studies.

Amplified fragment length polymorphism (AFLP) also utilizes variations within RE sites. The difference from RFLP is that, after the genomic DNA digestion with RE, the DNA fragments are ligated to adaptors that ligate to the restriction enzyme site. Primers complementary to adaptor sequences amplify a subset of ligated fragments. The amplified fragments are then electrophoresed, and the presence/absence of the fragments are visualized as variations for individuals. The advantage of AFLP is that it requires less amount of DNA compared to RFLP, and does not require radioactive probes. However, it is a dominant marker, so it is difficult to use as a marker in QTL analysis using segregating populations with a high rate of heterozygosity such as F_2 generation.

Random amplified polymorphic DNA (RAPD) markers are developed based on PCR amplification of random segments of genomic DNA with a single primer of the arbitrary nucleotide sequence. A disadvantage of RAPD markers is that it is a dominant marker, thus it cannot distinguish whether a locus is heterozygous or homozygous. Another disadvantage is that the amplification reproducibility is low because the amplification uses arbitrary primers that are not locus specific and the PCR result depends on genome DNA quality and PCR conditions.

3.2 Simple Sequence Repeats

Simple sequence repeats (SSR) markers or microsatellites are based on tandemly repeated 2–5 nucleotides, such as (CA)n, (AT)n, (ATT)n, and (ATG)n. Primers for SSR markers are designed using the conserved DNA sequences flanking these repeats. SSR is easy to use since it only requires PCR using the designated primers

to detect the repeat number variations between varieties. Repeat number variations can be observed as differences in fragment length using gel electrophoresis performed after PCR. SSR markers are abundant and easy to use, therefore they were the main markers used for QTL mapping prior to next-generation sequencing. Even now, SSR markers are still used to fill in the gaps in linkage maps produced from SNP genotyping, and it is useful for labs with limited budget and sometimes the genotyping using SSR is much faster and easy than SNP genotyping for a small-scale project or for fine-mapping a certain QTL locus.

A number of SSR markers for soybean has been developed since 1990s (Song et al. 2004). Song et al. (2004) evaluated the possibility of expressed sequence tag (EST) sequences as resource for SSR markers. The team screened 136,800 ESTs available in the GenBank to identify sequences containing SSRs. EST, however, contains low number of potential SSR markers having only low number of dinucleotide repeats of ten or more, and trinucleotide repeats of eight or more. Therefore, instead of EST, Song et al. (2004) looked into genomic libraries of 'Williams' soybean and bacterial artificial chromosome (BAC) clones. The number of SSR markers from genomic libraries and BAC clones was higher than that from EST. A total of 420 SSR markers developed: 24 from EST, and the remaining 396 markers were from BAC clones or genomic libraries. The SSR markers then were used to develop a genetic linkage map containing 20 linkage groups, with the number of markers per group varied between 12 and 29. After the release of Williams82 reference genome Glyma1, more than 210,000 potential SSR markers were identified from the reference genome (Song et al. 2010). After screening for locus specificity and perfect motif repeats, a BARCSOYSSR_01 database consisting of 21,206 SSR markers was published (Song et al. 2010). The database contains information on SSR marker ID, position in the chromosome, repeat motif, the physical position of flanking primers, estimated PCR product size, as well as the flanking primer sequences (Song et al. 2010).

3.3 Single-Nucleotide Polymorphism (SNP)

Single-nucleotide polymorphism (SNP) markers are abundant, biallelic, and their positions in the genome are known. In the GmHapMap dataset consisting of 1007 diverse accessions of soybean and wild soybean (Torkamaneh et al. 2021), there were 12.1 million SNPs across 1.1 Gb-length genomes. Biallelic means that at any SNP locus there will be only two nucleotides of four possible types (A, C, G, and T). The occurrence of the third allele is rare. The biallelic nature of SNPs also led to the development of automated and large-scale genotyping systems such as SNP chips or flexible array systems (Thomson 2014). Lastly, since the SNP position in the genome is known, it is possible to develop markers using primers based on the flanking sequences. These markers then can be used to compare genotyping results across different studies or populations and help the development of SNP databases compiled from different studies. In addition, the known position of SNPs makes it possible to predict the effect of SNPs on the phenotype. For example, SNPs located

in the exon may lead to an early stop codon, missense mutation, or induction of splicing variants. SNPs located in the promoter region may add or reduce cis-elements regulating the gene expression.

There are two types of SNP genotyping methods, fixed and flexible platforms (Thomson 2014). One example of a soybean fixed SNP genotyping platform is SoySNP50K BeadChip, which was developed based on DNA sequence analysis of six soybeans and two wild soybean accessions (Song et al. 2013). Initially, more than 200,000 SNPs were identified. After filtering for the distance between SNPs, unique flanking sequences indicating mapping to one locus, low rate of missing data, number of reads supporting each allele, and Illumina manufacturing phase, 52,041 SNPs distributed to 20 chromosomes were chosen as SoySNP50K content (Song et al. 2013). Genotype data of 12,116 soybean accessions has been used for genome-wide association mapping for oil and protein content (Bandillo et al. 2015). The GWAS study identified SNPs linked to oil and protein content on chromosomes 20 and 15, respectively (Bandillo et al. 2015). The advantages of a fixed SNP genotyping platform are high call rates because the SNPs included in the array already passed initial call rate filtering, fast turnaround time, cost-effectiveness per data point, the easiness to compare datasets resulting from different studies since the same set of SNPs are genotyped across samples, and, as a result, it is easier to create an SNP fingerprinting database (Thomson 2014). The main disadvantage of the SNP chip is the high cost to design a new SNP array. Therefore, a high-density SNP array is usually developed by company or institutions that are sure to use SNP array for a large number of samples. If the SNP array can be used as a "universal" array, meaning it can be used for genotyping a large number of samples, then the SNP chip will be cost-effective (Thomson 2014).

On the other hand, flexible SNP array platform examples are Douglas Array Tape, Fluidigm Dynamic Array, TaqMan, and KASP (Thomson 2014). The main characteristic of flexible array platforms is that these enable users to select markers for genotyping. Douglas Array Tape is a high-throughput platform having capacity to genotype more than 76,000 reactions per fun. Fluidigm Dynamic Array has less freedom in choosing number of markers and samples per run since it provides option either to genotype 96 SNPs × 96 samples or 24 SNPs × 192 samples. TaqMan and KASP can be run on a real-time PCR machine or fluorescence plate readers, thus the number of samples × markers in each run can be customized to fit in the 96-well or 384-well plate used in those machines. Region flanking target SNP is amplified using PCR and the SNP allele is detected using fluorescence tag attached to marker primers. If fixed array is more suitable for QTL mapping, GWAS, or genetic diversity analysis, flexible array is more suitable for marker-assisted selection (Thomson 2014).

Another type of a flexible SNP genotyping method is restriction enzyme (RE) sequence–based genotyping method. This method provides SNP data across the whole genome, thus it is suitable for QTL mapping or GWAS. There are several methods such as restriction site–associated DNA sequence (RAD-seq), double-digest RAD-seq (ddRAD-seq), and genotyping by sequencing (GBS). RE-sequence-based genotyping method does not require initial effort to develop SNP array but does require users to create barcoded adaptors to be linked to sample DNA fragments. In

addition, the sequencing cost is lower than normal deep whole genome sequencing because it combines multiple samples in one sequencing lane. Moreover, genomic library preparation can be performed in most labs equipped with molecular biology facilities. Therefore, users who do not have access to labs can do RE-sequence-based genotyping when they do not have SNP array. Disadvantages of RE-sequence-based genotyping is that it requires more computational resources for data analysis than fixed SNP array, the number and position of SNPs from genotyping can differ depending on the type of REs used in genomic library preparation, the SNP analysis pipeline.

As an example, GBS protocol will be described. The first step of the GBS protocol is the digestion of genomic DNA using two restriction enzymes (REs), a common-cutter RE and a rare-cutter RE (Poland et al. 2012). The digested DNA is ligated to adapters containing unique 4–6 bp nucleotides serving as a barcode sequence. A unique barcode sequence enables mixing samples between 96 and 384 samples into one sequencing tube (Poland et al. 2012). The sample mix is then sequenced using next-generation sequencing (NGS) machines. The resulting reads are then separated per sample based on barcode information, mapped to the reference genome, and called the variants. The tedious part of GBS is sequence read processing and data analysis, which needs computational skill for big data processing (Thomson 2014). Also, since multiple samples are mixed and sequenced in one sequencing reaction, the sequencing depth of each sample is shallow, which may produce a high missing rate in samples. Several bioinformatics tools are available for GBS data analysis and to impute (predict and filling-in) the missing SNP data, for example, TASSEL-GBS, IGST, Fast-GBS, UNEAK, or Stacks. The number of SNPs and accuracy of SNP calling can vary depending on the pipeline and parameters used.

3.4 Insertion-Deletion

Insertion-deletion (InDel) markers are developed based on small insertion-deletion (between 5 and 50 bp) when comparing two genome sequences. InDel can be detected using pipelines for SNP variant calling. Compared to SNP which is biallelic, insertion-deletion length can vary among individuals in diverse populations. Therefore, InDel marker genotype scoring cannot be easily automated and the markers are not commonly used for genome-wide association mapping. Nevertheless, the InDel marker is still useful in genotyping biparental segregating populations. In addition, InDel markers genotyping can be done in house, target fragments can be amplified by PCR, and variant detection can be done by gel electrophoresis – gel visualization.

4 QTL Mapping, GWAS, and Genomic Selection

4.1 QTL Mapping

Quantitative trait loci (QTL) analysis is a method to identify genomic regions associated with a quantitative trait by investigating phenotype-genotype association in a biparental segregating populations. Three types of populations are commonly

used in QTL mapping: F_2, recombinant inbred lines (RILs), and backcross inbred line (BIL) populations. F_2 population is derived from selfing the F_1 hybrid produced by crossing two parental lines having contrasting phenotype of target trait. The genome of F_2 lines is mostly heterozygous, therefore the F2 population is suitable for identifying major QTLs. RILs are developed by selfing F_2 individuals for 5–6 cycles, resulting in homozygous lines. Large number of seeds having identical genome composition can be produced from each RIL, therefore the RIL population is suitable for multi-years, multi-locations QTL analysis (Collard et al. 2005). Another type of population is BIL. BIL is developed by crossing back F_1 hybrids to one parental line (recurrent parent) for several cycles, followed by selfing until the genomes are fixed. BILs will have most of their genomes from the recurrent parent, and only a small part comes from the other parent. Since the genome is fixed, BILs can be used for multi-years and multi-locations studies. In addition, since the majority of the genome is from a recurrent parent, which is usually an improved line, promising BILs can be directly used in breeding programs.

The first step of QTL mapping is genotyping the population and constructing a linkage map or recombination bin map. Linkage map is constructed by calculating recombination frequency between markers, and the recombination frequency is translated to distance between markers (centiMorgan). One centiMorgan (cM) is defined as 1% chance of recombination between two markers. It means in a population of 100 individuals, there will be one recombination between two markers with 1 cM distance. In most of QTL analyses, marker distance calculation will involve more than two markers. There are two functions for marker distance calculation: Haldane mapping function, which assumes no interference between crossover, and Kosambi mapping function, which considers influence of recombination events to the adjacent recombination events (Collard et al. 2005). Recombination bin map approach is commonly used when genotyping is performed using GBS, RAD-seq, or SNP-chip. These genotyping methods produce high-density SNP map containing SNP markers that often are within the same haplotype, meaning they are closely located and segregated together. Using all SNPs in QTL mapping will require more computational resources and give redundant information. To reduce computational time and cost, only one SNP per bin is used in QTL mapping. This is already sufficient to represent the entire whole genome sequence (Patil et al. 2018). Representative SNPs are then used to create linkage map.

For QTL analysis, genotype, phenotype, and linkage map data are needed. There are several QTL mapping methods that are commonly used, such as single-marker analysis, interval mapping, or composite interval mapping (Collard et al. 2005). Single-marker analysis or single-marker association is finding association between one marker with phenotype. Single-marker analysis does not require linkage map data. Analysis of variance (ANOVA), linear regression, or *t*-tests are statistical methods used for single-marker analysis (Collard et al. 2005). Single-marker analysis can be performed without special QTL mapping software. However, QTL detection power will be less if the QTL is located far from the marker, due to recombination occurred between marker and QTL (Collard et al. 2005). Interval mapping incorporates marker distance information in QTL detection. While single-marker analysis only estimates correlation between QTL and one marker, interval

mapping can estimate QTL location between two markers (Collard et al. 2005). Composite interval mapping (CIM) is interval mapping method with linear regression and incorporates a subset of markers as covariates. By including covariates, the QTL analysis can account for linked QTLs and residual variation (Collard et al. 2005). CIM can detect QTLs with higher resolution but it requires more computational time.

After QTL is determined, the next step is fine mapping and candidate gene identification. This process largely changed after reference genomes and high-density SNP genotyping are available. First, the physical positions (position on genome) of flanking markers are determined. If the markers are generated by GBS or SNP chip, the marker physical position are already determined. Possible candidate genes are obtained by looking into reference genome sequence, for genes annotated within region between the flanking markers. Further selection can be done by checking whether the genes are expressed and the location or time of expression.

4.2 Genome-Wide Association Mapping (GWAS)

Genome-wide association mapping or genome-wide association studies (GWAS) utilize allelic diversity and recombination events in diverse populations to find genomic regions associated with certain trait (Chaudhary et al. 2015). One advantage of GWAS over QTL mapping is reduced time to create segregating populations needed for analysis. However, GWAS detection power relies on the frequency of minor allele in population and correct population structure (Chaudhary et al. 2015). Genotype data used for GWAS is usually filtered by removing SNPs having minor allele frequency less than 5% to avoid false positive, detect rare alleles (Chaudhary et al. 2015). GWAS studies have been conducted in soybean to identify SNPs associated with protein, oil, amino acid content, tocopherol, and isoflavones (Chaudhary et al. 2015; Wu et al. 2020; Bandillo et al. 2015; Sui et al. 2020). The challenge for GWAS analysis for nutritional components in soybean is the effect of environmental factors to the nutritional content. Diverse germplasm used in GWAS analysis may have different flowering time, thus temperature during seed filling period will differ for each accession. This may affect the phenotype and subsequently GWAS result.

4.3 Genomic Selection

Oil, protein, and isoflavone content are regulated by multiple and minor QTLs that interact with each other. Marker-assisted selection used a subset of markers linked to QTLs, therefore it may not correctly evaluate promising lines in breeding programs. Genomic selection term was introduced by Meuwissen et al. (2001). It utilizes all markers across the genome to predict phenotype, and this method is more affordable and efficient after the availability of NGS-based genotyping methods (Stewart-Brown et al. 2019). Genomic selection requires two populations, a training

population and test population. The training population will be genotyped and phenotyped, and both data are used to build a prediction model. Test population will be genotyped, and the genomic estimated breeding values (GEBV) will be estimated using prediction model built with training population. Possibility of using genomic selection in soybean was tested in Stewart-Brown et al. (2019), the protein and oil content were predicted using 483 elite breeding lines genotyped using BARCSoySNP6K iSelect BeadChips. The predictive ability for protein and oil was quite high, 0.81 and 0.71, respectively, compared with predictive ability for yield (0.26).

5 Genomic Resources and Other Bioinformatics Resources

5.1 Genome Assemblies and Reference Genomes

The estimated soybean genome size is 1.1-Gb, divided into 20 chromosomes (Schmutz et al. 2010). Approximately, 59 and 13 million years ago, genome duplications occurred twice, followed by several gene diversification and loss, as well as chromosome rearrangements (Schmutz et al. 2010). As a result, about 75% of soybean genes exist as multiple copies (Schmutz et al. 2010).

The first reference soybean genome was developed based on the Williams82 sequence (Schmutz et al. 2010), which was released as the Glyma1 version. The assembly was developed based on the whole-genome shotgun-sequencing method and it was one of the largest genomes to be assembled at that time (Schmutz et al. 2010). The assembly resulted in a total 950 Mbp length, grouped into 20 chromosomes and additional 1148 unanchored sequence scaffolds (Schmutz et al. 2010). The number of predicted protein-coding genes was 46,430 genes (Schmutz et al. 2010). About 78% of predicted genes were located in chromosome ends, where most all genetic recombination occurred (Schmutz et al. 2010). Genome sequence around centromeres is rich in repeats resulting in low recombination is this region (Schmutz et al. 2010).

The subsequent version (Wm82.a2.v1) replaced Glyma1 assembly by integrating a dense genetic linkage map produced from the RIL population of 'Williams 82' × wild soybean 'PI479752' and RIL population of 'Essex' × 'Williams82' (Song et al. 2016). This produced a genome assembly with a length of 949.2 Mbp for 20 pseudomolecules related to 20 chromosomes and additional 1170 unanchored scaffolds (Song et al. 2016), and 56,044 protein-coding genes (based on gene annotation file available in Phytozome.org). A major change in the Wm82.a2.v1 version is the change in the gene model name (Phytozome.org). Since some studies were conducted before the release of version 2.1, to know the relation between gene model in version 1 and version2.1, Soybase.org released a tool called Gene Model Correspondence Lookup where users can post the gene model either in version1 or version 2.1 and vice versa. As of June 2022, the current genome assembly of Williams82 has been improved to version Wm82.a4.v1. Wm82.a4.v1 was improved by incorporating the synteny of two other *Glycine* assemblies (Lee and Soja), and

also information based on PACBIO long reads. It has 961 Mbp of 20 pseudomolecules and 17 Mbp in unanchored scaffolds resulting in a total length of 282 scaffolds of 978.4 Mbp and the number of protein-coding genes is 52,872 genes (Valliyodan et al. 2019). The availability of multiple reference genomes for soybean will help identify genomic regions unique to certain varieties/species/regions.

5.2 Pangenomes

Pangenome represents the entire set of genes within a species, consisting of sequences that are shared among individuals or varieties in the species. The term pangenome was first defined in microbes (Tettelin et al. 2005). Core pangenome is entire genes shared by all strains within a clade (Tettelin et al. 2005). Shell pangenome is a gene set shared by several strains within a clade, whereas cloud pangenome is a gene set owned specifically by a strain within a clade. If the terms are translated to soybean, core pangenome can be defined as all genes shared by all varieties within soybean and wild soybean. Gene sets shared by several varieties in shell pangenome once were thought of as dispensable genes. However, recently it is thought that these genes may contribute to species diversity, encoding enzymes for a supplementary biochemical pathway that is not essential for growth but helps the species adaptation. Pangenome is getting attention instead of reference genome from one cultivar or variety because it can capture diversity present in the species.

The first pangenome study was published in 2014 (Li et al. 2014). Li et al. (2014) assembled de novo sequences of seven wild soybean accessions from Yellow River region, North and South China, Japan, Korea, Russia, and accession from the predicted domestication center in the China's northeast region. The de novo assemblies were compared to soybean reference genome Williams82 and identified 48.6% of the gene families as core genes (28,716 genes) (Li et al. 2014).

The second pangenome study included nine soybean landraces, 14 soybean, and three wild soybean accessions representing the diversity of 2898 deeply sequenced soybean and wild soybean accessions (Liu et al. 2020). De novo assembly of the 26 accessions produced genome assemblies with lengths varied from 992.3 Mb to 1059.8 Mb with average of 1011.6 Mb (Liu et al. 2020). The assemblies were then combined with Zhonghuang 13 as the primary reference genome into a graph-based pangenome (Liu et al. 2020). Orthologs analysis classified all genes from the 27 genomes into 57,492 families (Liu et al. 2020). Of these, 20,623 gene families were categorized as core genes since they were present in all 27 accessions (Liu et al. 2020). An addition of 8163 families were present in more than 90% of the collection (softcore genes), and the remaining 28,679 families were shell pangenome (Liu et al. 2020).

The third pangenome study (PanSoy) was published in 2021, involving de novo assemblies of 204 soybean accessions (Torkamaneh et al. 2021). The 204 accessions were selected as representative of the phylogenetic and geographical diversity of 1007 accessions from the GmHapMap dataset (Torkamaneh et al. 2021). PanSoy, sequence-based pangenome, was constructed by comparing each of 204 assemblies

to the 'Williams82' reference genome (Wm82.a4.v1; Valliyodan et al. 2019). PanSoy detected 54,531 gene families in the 204 accessions, 90.6% of which (49,431 genes) were identified as core genes (Torkamaneh et al. 2021). The PanSoy genome coverage was evaluated using 'Lee' reference genome (Valliyodan et al. 2019). About 99.9% of 'Lee' genome sequence could be mapped to the PanSoy genome, in comparison to 92% of the 'Lee' genome mapped to the Williams82 reference genome only (Torkamaneh et al. 2021). The fourth pangenome study was based on resequencing data of 1110 accessions from the USDA Soybean Germplasm Collection (Bayer et al. 2021). The analysis produced a pangenome with 1213 Mbp length and 51,414 predicted gene families (Bayer et al. 2021). Of more than 50,000 gene families, 86.8% of genes are core genes (Bayer et al. 2021).

5.3 Databases and Resources for Genetic Research

Online databases for reference genome sequences, gene annotation, resequencing data, SNPs, transposable elements, biosynthesis pathway, transcriptome, proteome, metabolomics, cyst nematode proteins, functional network, functional genomics, and root phenotype have been developed for soybean. These databases contain data that can help other researchers in furthering their research, however, unfortunately, many of these databases were not accessible anymore or were not easy to be found. Two databases, Soybase (www.soybase.org, last accessed April 5th, 2023) and Phytozome (www.phytozome.org, last accessed April 5th, 2023), will be introduced here.

Soybase (Grant et al. 2009) was developed by USDA-ARS. It provides a comprehensive repository of professionally curated genetics, genomics, and other related data for soybean analysis (Grant et al. 2009). Currently, it houses a large variety of data, from genetic marker data, genetic map, a compilation of QTLs and GWAS peaks identified in past, transcriptome data, SNP data from resequencing projects, whole-genome sequence data, and pangenomes to gene ontology, and tutorials on soybean development, disease, and pests. Transcriptome datasets from Severin et al. (2010) and Le et al. (2007) are also provided here.

Severin et al. (2010) provided transcriptome data from 14 soybean tissues (leaf, flower, one M-pod, 10 days after flowering (DAF), and 14-DAF pods, root, nodule, and seven stages of developing seeds). The data is in Glyma1, thus users need to convert the gene IDs to Wm82.a2.v1 version and vice versa. On the other hand, Gene Networks in Seed Development project (Le et al. 2007) is based on transcriptome profiling of soybean seed regions (seed coat, endosperm, and embryo subregions). It is also known as Goldberg/Harada dataset. It used laser capture microdissection (LCM) to isolate these subregions. A total of 40 soybean seed sections were analyzed. Detailed analysis using seed sections was performed at the following stages: globular stage, heart stage, cotyledon stage, and early maturation stage. In addition, data from the mid-maturation stage, late-maturation stage, dry seeds, trifoliate leaves, roots, stems, floral buds, and seedlings 6 days after imbibition are also provided to represent the soybean life cycle. The datasets provide

knowledge of genes active in different seed parts during development, transcription factors localized at specific seed regions and subregions, and biological processes important for seed differentiation. The dataset can be accessed from http://seedgenenetwork.net (last accessed July 7th, 2022) and from the Soybase Expression Explorer Database. There are also transcriptome datasets grouped by experiments and are provided under NCBI GEO Expression data.

Phytozome (Goodstein et al. 2012) is developed and is maintained by the Department of Energy, Joint Genome Institute. It serves as a plant comparative genome portal and the current version (Phytozome v13) houses 304 assembled and annotated genomes (last accessed 4th April 2023). Phytozome v13 hosts three soybean reference genomes, Williams82, Lee, Fiskeby III, and one wild soybean reference genome PI 483463. Users can perform BLAST analysis, search genes using keywords or gene ID, or retrieve sequences based on genomic position. For soybean and wild soybean, users can obtain the gene genomic sequence, transcript, coding sequence, and peptide sequence. Information about its function, organs where it is expressed, KEGG pathway annotation, gene homologs, and similarity to the homologs in other species are also available. Users can also perform gene comparative analysis across different species. Registered users can download whole-genome sequence (all softmasked, hardmasked, and repeat masked), transcript data, and gene annotation.

6 Soy-Based Food

Soy-based food has played an important contribution to the health, culture, and economy of Asian countries, especially China, Japan, Korea, and Indonesia. It is a high-protein food and contains essential fatty acids such as linoleic acid and linolenic acid. Soybean is known as “meat in the field” since it supplies high level of protein. It also provides nutritional components that provide health benefits beyond nutrients and other non-nutritional components with physiological function. There are currently three groups of Food for Specified Health Uses (FOSHU) soy-based products: cholesterol lowering, intestinal function regulation, and bone health (Yamamoto 2005). The versatility of soy for human consumption is shown in the rich varieties of soy-based food products.

The protein and oil in soybeans are the components that are mostly made use of. The protein content in soybeans is an important aspect to know because they are often the ones that will experience the most changes when being processed. During fermentation of several traditional soy-based products, such as tempeh and miso, the protein content is metabolized to amino acids which will result in higher nutritional value and also enhance its texture and taste (Handoyo and Morita 2006). Notable amino acids are the umami-inducing glutamate and arginine. Soybean and products made of soybeans with high glutamic acid content often tastes better and is more preferred. When the soybean is further processed, like in tempeh, for example, the glutamic acid content increases as fermentation progresses (Gunawan-Puteri et al. 2015). The glutamic acid content gives tempeh its umami taste. Arginine on the other

hand has an important function in metabolism reactions like, for example, in the urea cycle (Morris 2002). It helps in synthesis of larger protein molecules in the body. In the further processing of soybeans, arginine is one of the primary amino acids that could be used, for example, in fermentation.

Soybean is also a very promising ingredient for functional food products. Soybean contains significant amounts of bioactive components, mainly phenolic compounds like flavonoids and isoflavones, that have antioxidant activities which are found to be enhanced in soybean products. Alongside protein and other nutrients in soybean, the bioactive components give soybean its functional properties such as antidiabetic, anticancer, anti-inflammatory, estrogenic, and anti-hypercholesterolemic function.

Soy anti-nutrients, such as phytic acid, are eliminated with fermentation and soaking in processed soy foods such as tempeh, soy sauce, and tauco (Hui et al. 2004). Soybean could be processed using fermentation or without the fermentation process. Some of the soybean products that are fermented using *Lactobacillus bulgaricus* and *Mucor sufu* are soy yogurt and soy cheese, respectively. Well-known fermented products in Indonesia are tempeh, soy sauce, and tauco. Tempeh is fermented using a mould starter of *Rhizopus oligosporus*, soy sauce is made by inoculating with *Aspergillus sojae* mould, while miso is fermented using *Aspergillus oryzae* mould. Natto is another soybean fermented food with *Bacillus subtilis*. Examples of soy-based foods that do not undergo fermentation process are tofu or soymilk. Soymilk can be further processed into tofu. Nonfermented soy products processed with modern methods include oil, soy flour, and soy protein isolate. Soy protein could also be used as an ingredient to produce synthetic meat (TVP/textured vegetable protein). Learning about the soybean ingredient requirement for each food product might become the future in soybean cultivation development.

6.1 Vegetable Soybean

Vegetable soybeans are defined as those that are harvested after the R6, the immature green state, but before the R7, the beginning of maturity stage when pods located at the main stem start to change into brown or tanned color. Vegetable soybean has different names: edamame (Japan), poot kong (Korea), maodou (China), and tuarae (Thailand). Edamame – green pods attached to the stem – is a popular snack in Japan. It is generally sold in the pods as fresh or frozen beans, attached or detached from the stem, though canned and frozen shelled beans are also available in the market. For consumption, edamame is boiled for 5–7 min in highly salted water. However, it can also be shelled, removed from the pods for boiling, baking, steaming, or toasting over fire for making soup, salad, or vegetable dishes. And while it is logical that vegetable soybeans that have 9–10 times moisture content than the mature bean have lower other nutrient components, surprisingly vegetable soybeans are recorded to have higher vitamin A, vitamin C, sodium, and comparable niacin content. In the dry weight composition, vegetable soybean also contains more calcium, folate, and comparable protein, fat, ash, carbohydrate, and fiber content compared to the mature soybean seed.

The sensory quality of vegetable soybean owes to its sweetness and umami tastes along with the texture and aroma quality. Vegetable soybeans contain 1.28–7.12 g of sucrose, 0.37–1.51 aspartic acid, 0.64–2.82 glutamic acid, 0.17–0.72 glycine, and 0.11–0.51 g alanin per 100 g of bean (Kumar et al. 2011). The sodium content of vegetable soybean is higher than mature soybean (15 and 2 mg/100 bean), explaining the higher salty umami interaction of the sodium with umami inducing amino acid. It has significantly higher moisture content (67.5%) compared to mature soybean (8.54%) and was valued to have higher flavor quality compared to its mature soybean.

The morphological characteristics expected for a good quality vegetable soybean include pod size, number of beans per pod, and pod color (Shanmugasundaram and Yan 2010). Preferable traits for vegetable soybean are large seeds (>30 g per 100 g dry seeds), with two or more beans per pod, subsequently pod size (should be ≥5.0 cm and ≥1.4 cm) as well as number of pods (should be ≤175 pods per 500 g frozen pod packet) become a quality parameter in vegetable soybean. The preferable pod and bean color for vegetable soybean is dark green, however, the mature seed color can be diverse from yellow, green, brown/red, or black.

6.2 Tempeh

Soy tempeh is defined as a compact white cake that is prepared from dehulled, boiled, and acidified soybeans through solid-state fermentation with *Rhizopus* spp., originated from Indonesia. Tempeh uses the whole bean of mature soybean and thus creates a hearty and firm, meaty texture from mycelium with a grizzle of soft bean texture. It has savory, nutty, earthy, and mushroom flavors.

Tempeh is unique among major traditional soy foods because it did not originate from China or Japan. Variation occurs among tempeh artisans in the dehulling, boiling, and acidification processes prior tempeh starters inoculation. The acidification process can be conducted naturally by overnight soaking or chemically by soaking the bean in food-grade acidulants. Once the acidified soybean is inoculated with the mold, it is common to directly pack them into porous plastic bags or banana leaves and leave them to begin the solid-state fermentation. Tempeh can be consumed as it is, but the most popular preparation is to fry, stir-fry, or boil them with salt and spices and/or flour. The popularity of tempeh within the nation was never in question, as this relatively inexpensive meat replacement for protein source has been consumed by millions of Indonesian daily.

Tempeh has high socioeconomic importance in Indonesia as there are more than 115,000 micro-, small, and medium enterprises (MSMEs), which employ 285,000 workers and generate about 57 million EUR per year. The soybean tempeh industry alone absorbs more than 1.2 million tons of soybeans per year (Central of Indonesian Agricultural Data and Information System 2013), which is more than 70% of soybean consumption within the country. The importance of the tempeh industry in Indonesia has affected domestic soybean development as it also put importance on

the tempeh requirement such as the seed size, in addition to the yield, harvest time, and resistance to abiotic stress.

According to the standard quality of setup by National Standardization Agency of Indonesia, soy tempeh contains minimum 15% protein and 7% fat (National Standardization Agency of Indonesia 2015), while the USDA National Nutrient Database recorded that tempeh has approximately 20% protein and 10% fat. Several studies recorded that Indonesian domestic soybean in general is having higher protein and lower fat content compared to imported soybean, especially those from the USA. The main reason is that soybean is used in the USA for soybean oil production. Therefore, the soybean demand is more a derivative of animal product consumption rather than for direct human consumption, such as tempeh and tofu in Indonesia.

The first step in tempeh production is soybean hydration by soaking or boiling in water. This process also aims to loosen up the hull and facilitate the wet dehulling process. Traditionally, wet dehulling processes in tempeh are done manually by human feet as the workers get into the large pool of soybean and step on them. Nowadays, due to hygiene reasons and technological availability, dehulling machines have been employed for the process, and some more advanced tempeh industries are even employing dry dehulling techniques prior to water hydration. As tropical countries with high humidity and room temperature, microorganisms are naturally attracted during soaking which leads to natural lactic acid bacteria (LAB) acidification. The acidification is an important step in ensuring the success of mould inoculation as it wipes out other competitor microorganisms and provides an ideal condition for *Rhizopus* moulds to start the solid-state fermentation. In temperate countries where the humidity and temperature are much less, chemical acidification is often added to support this process.

After the acidification process, the bean is boiled, drained, and cooled prior to *Rhizopus* mold inoculation. Traditionally, the mold is introduced by mixing small amounts of tempeh from previous fermentation, but nowadays ready-to-use and optimized tempeh starters have been commercially available. Following the tempeh starter inoculation, the inoculated beans are wrapped with clean banana leaves or porous plastic bags and allowed to sit at room temperature for at least 24 h. By this time, mycelium hold the beans together, though the white mycelium might not be visually available yet. Tempeh producers usually have this stage of tempeh (tempepera) during distribution, allowing more fermentation to happen before it reaches the customers. The fresh ripe tempeh covered with white mycelium is usually consumed within 2 days, otherwise it goes into the next stage of overripe tempeh. While fresh ripe tempeh is often consumed as a whole meal, overripe tempeh has much stronger taste and aroma and is often used more as condiment.

Tempeh is a traditional Indonesian fermented food known for its high nutrients and superior digestibility (Hermana and Karyadi 1997). Tempeh fermentation increases sodium content that may contribute to salty taste and umami interaction with free amino acids. Tempeh also has a higher composition of manganese, niacin, and vitamin B-6 compared to soybean. Interestingly, tempeh contains vitamin B12 that commonly came from animal source food and was not available in raw mature soybean.

The protein and fat content per 100 g dry weight of tempeh is higher than per 100 g dry weight of soybean. Tempeh fermentation increases protein digestibility and free amino acid content, and the availability of other macronutrients through hydrolysis and reduction anti-nutrient such as phytate, saponin, and protease inhibitors, making it an even better plant-based protein source than soybean. The nutrient content and availability in the originated soybean remain important as it will support the microbial growth.

Tempeh consumption is associated with cholesterol reduction in blood and helps prevent cardiovascular diseases (Kris-Etherton et al. 1999; Mangkuwidjojo et al. 1985). It also shows diarrhea prevention effect, which improves its role in food safety area (Kiers et al. 2003). Tempeh has also found its way in the global market with the increasing findings of its benefits and the growing lifestyle toward plant-based food. At the current time, tempeh has been produced commercially in Japan, India, the USA, Canada, European countries, Australia, and New Zealand.

6.3 Natto

Natto is produced through fermentation of soybean by *Bacillus subtilis*. It originates and an important commodity in Japan. Natto is not the only soy-based *Bacillus* fermented product. In other countries, *Bacillus* fermented soy food has also been traditionally integrated into their national culture, such as chongkukjang in Korea, kinema in India, and thuanao in Thailand.

Natto has a dark color, pungent but pleasant aroma, and sticky viscous coating with cheese texture. It can be served as it is or used as a seasoning agent with raw or cooked seafood, meat, and vegetables. Natto contains higher vitamin C content than soybean seed and has approximately 43, 24, 12, and 10% dry weight of protein, fat, fiber, and sugar content, which is a higher proportion compared to soybean. The proportion (but not the content) of several minerals, such as calcium, iron, sodium, zinc, manganese, and selenium, in natto are also higher than those in soybean. The peptide, free amino acids, ammonia, saccharide hydrolysis products, and minerals produced during fermentation constitute the characteristic flavor of natto. The sticky substances covering the well-fermented natto are composed of polyglutamic acid and levan (fructan) (Claus 1986).

The *B. subtilis* fermentation also produces proteases and amylase that help with other food digestion processes inside the human body (Ferrari et al. 1993). Serine protease of subtilisin is shown to degrade Gly m Bd 28K (Ogawa 2000). Serine protease is also recorded to show fibrinolytic activity (Sumi et al. 1990), the formation of blood clots from platelets and blood-derived proteins (fibrin), which prevents extended bleeding and promote healing. Both catalase and subtilisin produced during natto fermentation exhibit a growth-promoting effect on fecal microflora (Terada et al. 1999), which adds more to its hypoallergenic character (Kalliomaki et al. 2001), as well as providing positive impact to gut health (Hosoi et al. 2000) and immune function (Pelto et al. 1998).

The natto quality is affected by the soybean quality and *B. subtilis* (natto) strains. Japanese domestic soybeans such as ‘Suzuhime’, ‘Suzumaru’, ‘Kosuzu’,

and 'Natto-shoryu' are preferred for natto production because of their small seeds (5.5–7.3 mm) or extra small (< 5.5 mm) in size, although there are natto produced from midsize or large-seeded soybeans. The seed size was reported to affect negatively with the activities of subtilisin alkaline protease that play roles in protein degradation and intensity of the taste and smell of natto. Therefore, natto made from larger seed soybean have weaker smell while those from extra small seed may have excess fermentation and stronger taste (Takahashi et al. 1996). Natto producers prefer soybean with yellow seed due to unappetizing appearance and lack of aroma character of brown and black soybean. It is just recently that black soybean natto is being on the shelf with highlights on the higher content of polyphenols.

Other than its morphological characters, the soybean contents also play an important role in producing high-quality natto. Higher sugar content is associated with better taste and flavor production as it also has a positive correlation with the bacteria growth as *B. subtilis* natto strain can utilize di- and oligosaccharides but not starch and fiber. Smaller seed size soybean varieties are also associated with higher sugar content and therefore preferable for natto production. The protein content and profile of the originated soybean also affect the natto quality as *B. subtilis* natto strain utilize protein, peptides, and amino acids for their growth. Soybeans with high protein content are preferred. Among Japanese domestic soybean used for natto production, 'Suzuhime' and 'Zizuka' cultivars are highly regarded starting ingredient as they have high protein content, bright color, and polished appearance (Hosoi and Kiuchi 2003).

In natto processing, soybeans are washed, soaked, and steamed prior to the *B. subtilis* (natto) spore inoculation. The inoculated beans are then packed and left for solid-state fermentation. Overnight soaking in cold water and steaming of the beans facilitates water hydration, bean swelling, and denature undesirable protein. Steaming also removes contaminant and pathogenic bacteria. Right after steaming, while the beans are still hot, *B. subtilis* (natto) spore suspension is sprayed onto the bean. The fermentation is set on 40 °C with 85–90% humidity whereas the humidity is reduced to 75 and 55% in the following 6 and 16 h after fermentation started. Traditionally the inoculated beans were packed in rice straw to maintain warmth and absorb emitted carbon dioxide during the initial stage of fermentation.

6.4 Miso

Miso is a traditional Japanese fermented soy paste with thick texture, salty taste, savory flavor and aroma, and white, red, or brown color. It has high protein and low-fat content. Miso is often used as a soup base though it may also be used for meat and seafood seasoning prior cooking, or as condiment in many traditional Japanese dishes.

There are two steps in miso production. First is koji production. In this process, *Aspergillus* fungi is inoculated to growth materials. The most common growth material is rice, although koji can also be produced using barley, or soybeans (Shurtleff and Aoyagi 2018). Second is the fermentation of boiled soybean by salt

and koji. In this step, koji interacts with salt and moisture to induce the yeast and bacteria growth of miso. Miso fermentation period is 1 week to 3 years, and the color and flavor intensity increase with longer fermentation time (Ogasawara et al. 2006). According to the amount of salt and length of fermentation, miso can be classified into white, red, or dark miso (Katz 2012). White miso uses more koji, less salt, and applies shorter fermentation time, while dark miso is the opposite, and red miso is in between the arrangement of white and dark miso.

Traditionally, miso was a homemade but commercial production is a major form since 1980. About 90% of miso produced in Japan came from factories (Shurtleff and Aoyagi 2018). The commercial miso production was boosted by mechanization, the availability and usage of commercial strains, standardization of processing methods, and the improvement of packaging and sterilization process (Shurtleff and Aoyagi 2018).

Not all soybean can meet the quality parameters of soybean suitable for miso production. And of them that are suitable, most of them are Japanese domestic soybeans. The requirements for soybean to be selected for miso production include white hilum color, light yellow color of cotyledon, a high water-absorbing capacity during soaking, softness, sucrose content, and protein content. Though the color requirements might slightly change due to the popularity of high polyphenol of the darker color soybean, in general light yellow soybeans were assumed to give a more appetizing end product in the miso production. As miso end product is in the form of paste, the soft structure and the high water absorbing capacity is important to ensure targeted tenderness after soybean cooking. The water absorbing capacity also correlates with higher carbohydrates that gives even more free sugars. Other than influencing the taste of the end product, free sugars along with free amino acids availability prior inoculation are known to support the microbial growth during miso fermentation.

Soy-fermented products similar to miso include tauco from Indonesia and also gochujang and doenjang from Korea. The tauco production differs from miso in terms of soybean form and the mold employed. The soybean preparation is similar without the mashing process. The soaked and cooked bean is fermented with salt and tempeh starter or combination of *Rhizopus oligosporus*, *R. oryzae*, and *A. oryzae*. Both gochujang and doenjang are fermented soybean paste that employ *Aspergillus* mold and brine fermentation. The soybean preparation is similar to miso with an additional process of square forming the mashed cooked soybean and drying them into large bricks. Inoculation of *Aspergillus* is done directly to the dried soybeans for 20–90 days resulting in fermented soybean bricks, called meju. Meju is the intermediate product that will be processed into either gochujang, or ganjang and doenjang, the trinity of Korean traditional condiments (Patra et al. 2016). In traditional production, Meju is broken down into small pieces and put into onggi, Korean traditional earthenware for storage or fermentation. To create gochujang, the meju is mixed with red chili powder, glutinous rice powder, salt, garlic, and onion, and then sweetened with a little sugar syrup and aged in onggi, a type of Korean earthenware (Song et al. 2021). To create ganjang and doenjang, brine is added to the broken meju and salt fermentation inside the onggi was performed for about 2 months creating the liquid part (kanjang) and the solid part (doenjang) (Patra et al. 2016).

6.5 Third-Generation Product from Soy Processing

In the first generation of soy processing, whole soybean is being used, for example, in the production of tempeh and natto. The second-generation soy products refer to those coming from part of the soybean such as soybean oil, soy milk, and soy sauce. The third-generation products are targeted compounds that are isolated or fractionated from soy. The third-generation products are commonly components with high benefits for specific health uses which are required in higher amounts than those that can be acquired by only consumption. The production commercially of third-generation products require solid scientific evidence of the component benefit, as well as advance investment in technology and therefore the products are commonly having a high economic value. Up to this book is being written the following third-generation products from soy processing have been commercialized either as supplement or food ingredients: isolate soy protein, soy peptides, soy isoflavone/phytoestrogen, soy lecithin, soy fiber, soy phytosterol, and soy oligosaccharides.

Protein and bioactive peptide contents of soybean contribute to its potential to prevent lifestyle-related diseases (Yoshikawa and Tsuruki 2005). Soy peptides have been reported to have hypocholesterolemic (Sugano et al. 1988), hypotriglyceridemic, hypotensive, anticancer, and immunomodulating activities while also having an impact on the regulation of food intake. While the roles of dietary peptide in nutrition have been well established, soy peptides have been used in the formulation of functional foods (Takamatsu 2005) to relieve sports fatigue and stress, overcoming obesity, lowering blood cholesterol, hypoallergenic food, and seemed to have a future in alleviating stress and improvement of brain function. Hypocholesterolemic nature of soy protein comes from its hydrophobic high-molecular-weight peptides left after digestion that binds to bile acids (Carroll 1991). The bile acid–binding ability of soy peptide increased fecal excretion of bile acids and reduced cholesterol level in serum and liver. Soy peptide was also reported to stimulate fat metabolism and suppress fatty acid synthesis, leading to its hypotriglyceridemic activity.

7 Conclusion

Numerous studies have been conducted on nutritional content and physiological benefits beyond nutrition of soybean seeds as well as the regulation of biosynthesis of these components. Knowledge on nutritional values and bioactive content of soybean seeds can be utilized to increase their content in current soybean varieties through breeding. The availability of next-generation sequencing technologies, high-density genotyping platforms, and soybean reference genomes have accelerated the elucidation of genetic basic of the biosynthesis of these compounds. The knowledge of genetic basis of these traits enables breeders to select potential breeding lines efficiently using DNA markers. The breeding lines then can be utilized to produce soybean-based food products that are not only nutritious but also meet the diverse needs of consumers and the food industry.

References

Adachi M, Kanamori J, Masuda T et al (2003) Crystal structure of soybean 11S globulin: glycinin A3B4 homohexamer. PNAS 100(12):7395–7400. https://doi.org/10.1073/pnas.0832158100

Anai T, Yamada T, Hideshima R et al (2008) Two high-oleic-acid soybean mutants, M23 and KK21, have disrupted microsomal omega-6 fatty acid desaturase, encoded by GmFAD2-1a. Breed Sci 58:447–452. https://doi.org/10.1270/jsbbs.58.447

Bandillo N, Jarquin D, Song Q et al (2015) A population structure and genome-wide association analysis on the USDA soybean germplasm collection. Plant Genome. https://doi.org/10.3835/plantgenome2015.04.0024

Bando N, Tsuji H, Hiemori M et al (1998) Quantitative analysis of Gly m Bd 28K in soybean products by a sandwich enzyme-linked immunosorbent assay. J Nutr Sci Vitaminol 44:655–664

Bayer PE, Valliyodan B, Hu H et al (2021) Sequencing the USDA core soybean collection reveals gene loss during domestication and breeding. Plant Genome 15:1–12. https://doi.org/10.1002/tpg2.20109

Bilyeu K, Ren C, Nguyen HT et al (2009) Association of a Four-Basepair Insertion in the P34 gene with the low-allergen trait in soybean. Plant Genome 2:141–148. https://doi.org/10.3835/plantgenome2009.01.0006

Britz SJ, Kremer DF (2002) Warm temperatures or drought during seed maturation increase free α-tocopherol in seeds of soybean (*Glycine max* [L.] Merr.). J Agric Food Chem 50:6058–6063. https://doi.org/10.1021/jf0200016

Carrão-Panizzi E (2007) Environmental and genetic variation of soybean tocopherol content under Brazilian growing conditions. J Am Oil Chem Soc 10:921

Carroll KK (1991) Review of clinical studies on cholesterol-lowering response to soy protein. J Am Diet Assoc 91(7):820–828

Central of Indonesian Agricultural Data and Information System (Pusat Data dan Sistem Informasi Pertanian) (2013) Soybean (Original title: Kedelai). Buletin Konsumsi Pangan 4(3):8–16

Chaudhary J, Patil GB, Sonah H et al (2015) Expanding omics resources for improvement of soybean seed composition traits. Front Plant Sci 6:1–16. https://doi.org/10.3389/fpls.2015.01021

Chennupati P, Seguin P, Liu W (2011) Effects of high temperature stress at different development stages on soybean isoflavone and tocopherol concentrations. J Agric Food Chem 59:13081–13088. https://doi.org/10.1021/jf2037714

Claus D, Berkeley RCW (1986) Genus bacillus subtilis Cohn, 1872, 174. In: Sneath PHA, Mair NS, Sharpe ME, Holt JG (eds) Bergey's manual of determinative bacteriology, vol. 2. Williams & Willkins, Baltimore, pp 1105–1139

Clemente TE, Cahoon EB (2009) Soybean oil: genetic approaches for modification of functionality and total content. Plant Physiol 151:1030–1040. https://doi.org/10.1104/pp.109.146282

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142:169–196. https://doi.org/10.1007/s10681-005-1681-5

Demorest ZL, Coffman A, Baltes NJ et al (2016) Direct stacking of sequence-specific nuclease-induced mutations to produce high oleic and low linolenic soybean oil. BMC Plant Biol 16:1–8. https://doi.org/10.1186/s12870-016-0906-1

Dhaubhadel S, McGarvey BD, Williams R et al (2003) Isoflavonoid biosynthesis and accumulation in developing soybean seeds. Plant Mol Biol 53:733–743. https://doi.org/10.1023/B:PLAN.0000023666.30358.ae

Dhaubhadel S, Gijzen M, Moy P, Farhangkhoee M (2007) Transcriptome analysis reveals a critical role of *CHS7* and *CHS8* genes for isoflavonoid synthesis in soybean seeds. Plant Physiol 143:326–338. https://doi.org/10.1104/pp.106.086306

Dwiyanti MS, Maruyama S, Hirono M, et al. (2016) Natural diversity of seed α-tocopherol ratio in wild soybean (Glycine soja) germplasm collection. Breed Sci 66:653–657. https://doi.org/10.1270/jsbbs.16028

Dwiyanti MS, Yamada T, Sato M et al (2011) Genetic variation of γ-tocopherol methyltransferase gene contributes to elevated α-tocopherol content in soybean seeds. BMC Plant Biol. https://doi.org/10.1186/1471-2229-11-152

Ferrari E, Jarnagin AS, Schmidt BF (1993) Commercial production of extracellular enzymes. In: Sonenshein AL, Hoch JA, Losick R (eds) Bacillus subtilis and other gram-positive bacteria. American Society for Microbiology, Washington, DC, pp 917–937

Goodstein DM, Shu S, Howson R et al (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:1178–1186. https://doi.org/10.1093/nar/gkr944

Grant D, Nelson RT, Cannon SB, Shoemaker RC (2009) SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Res 38:843–846. https://doi.org/10.1093/nar/gkp798

Gunawan-Puteri MDPT, Hassanein TR, Prabawati EK (2015) Sensory characteristics of seasoning powders from overripe tempeh, a solid state fermented soybean. Procedia Chem 14:263–269

Handoyo T, Morita N (2006) Structural and functional properties of fermented soybean (Tempeh) by using *Rhizopus oligosporus*. Int J Food Prop 9(2):347–355. https://doi.org/10.1080/10942910500224746

Hermana MM, Karyadi D (1997) Composition and nutritional value of Tempe: its role in the improvement of the nutritional value of food. In: Agranoff J (ed) The complete handbook of Tempe: the unique fermented Soyfood of Indonesia. American Soybean Association, Singapore

Hosoi T, Kiuchi K (2003) Natto–a food made by fermenting cooked soybeans with Bacillus subtilis (natto). In: Handbook of fermented functional foods. CRC Press, Boca Raton, pp 227–245

Hosoi T, Ametani A, Kiuchi K et al (2000) Improved growth and viability of lactobacilli in the presence of *B. subtilis* (natto), catalase, or Subtilisin. Can J Microbiol 46:892–897

Hui YH, Meunier-Goddik L, Josephsen J et al (eds) (2004) Handbook of food and beverage fermentation technology, vol 134. CRC Press

Jiang Q, Christen S, Shigenaga MK, Ames BN (2001) γ-Tocopherol, the major form of vitamin E in the US diet, deserves more attention. Am J Clin Nutr 74:714–722. https://doi.org/10.1093/ajcn/74.6.714

Kalliomaki M, Salminen S, Arvilommi H (2001) Probiotics in primary prevention of atopic disease: a randomized placebo controlled trial. Lancet 357:1076–1079

Karunanandaa B, Qi Q, Hao M et al (2005) Metabolically engineered oilseed crops with enhanced seed tocopherol. Metab Eng 7:384–400. https://doi.org/10.1016/j.ymben.2005.05.005

Kato S, Yumoto S, Takada Y et al (2007) A new soybean cultivar 'Kinusayaka' lacking three lipoxygenase isozymes and group A acetyl saponin. Bull Natl Agric Res Cent Tohoku Reg 107:29–42

Katz SE (2012) The art of fermentation. Chelsea Green Publishing

Kiers JL, Nout MJR, Rombouts FM et al (2003) Effect of fermented soya beans on diarrhoea and feed efficiency in weaned piglets. J Appl Microbiol 95:545–552

Kitamura K (1995) Genetic improvement of nutritional and food processing quality in soybean. Jpn Agric Res Q 29:1–8. https://doi.org/10.6013/jbrewsocjapan1988.89.926

Kofsky J, Zhang H, Song BH (2018) The untapped genetic reservoir: the past, current, and future applications of the wild soybean (*Glycine soja*). Front Plant Sci. https://doi.org/10.3389/fpls.2018.00949

Kris-Etherton PM, Yu-Poth S, Sabaté J et al (1999) Nuts and their bioactive constituents: effects on serum lipids and other factors that affect disease risk. Am J Clin Nutr 70:504S–511S

Krishnamurthy P, Min J (2014) Evaluation of genetic structure of Korean wild soybean (*Glycine soja*) based on saponin allele polymorphism. Genet Resour Crop Evol 61:1121–1130. https://doi.org/10.1007/s10722-014-0095-4

Kumar V, Rani A, Goyal L et al (2011) Evaluation of vegetable-type soybean for sucrose, taste-related amino acids, and isoflavones contents. Int J Food Prop 14(5):1142–1151

Le BH, Wagmaister JA, Kawashima T, Bui AQ, Harada JJ, Goldberg RB (2007) Using genomics to study legume development. Plant Physiol 144:562–584. https://doi.org/10.1104/pp.107.100362

Li YH, Zhou G, Ma J et al (2014) De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nat Biotechnol 32:1045–1052. https://doi.org/10.1038/nbt.2979

Liu Y, Du H, Li P et al (2020) Pan-genome of wild and cultivated soybeans. Cell 182:162–176.e13. https://doi.org/10.1016/j.cell.2020.05.023

Mangkuwidjojo S, Pranowo D, Nitisuwirjo S et al (1985) Observation on hypolipidemic activity of Tempe. In: Hermana Karyadi D (ed) Proceedings of the symposium on the utilization of Tempe in health and nutrition development. Nutrition Research and Development Center, Department of Health of Republic of Indonesia, Bogor, pp 114–127

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Morris SM Jr (2002) Regulation of enzymes of the urea cycle and arginine metabolism. Annu Rev Nutr 22(1):87–105

National Standardization Agency of Indonesia (Badan Standardisasi Nasional) (2015) Soy Tempe (Original title: Tempe Kedelai) SNI 3144:2015. Badan Standardisasi Nasional, Jakarta

Nout MJR, Kiers JL (2005) Tempe fermentation, innovation and functionality: update into the third millenium. J Appl Microbiol 98:789–805. https://doi.org/10.1111/j.1365-2672.2004.02471.x

Ogasawara M, Yamada Y, Egi M (2006) Taste enhancer from the long-term ripening of miso (soybean paste). Food Chem 99(4):736–741

Ogawa A, Samoto M, Takahashi K (2000) Soybean allergen and hypoallergenic soybean products. J Nutr Sci Vitaminol 46:271–279

Park C, Dwiyanti MS, Nagano AJ et al (2019) Identification of quantitative trait loci for increased α-tocopherol biosynthesis in wild soybean using a high-density genetic map. BMC Plant Biol 19:1–15. https://doi.org/10.1186/s12870-019-2117-z

Patil G, Mian R, Vuong T et al (2017) Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. Theor Appl Genet 130:1975–1991. https://doi.org/10.1007/s00122-017-2955-8

Patil G, Vuong TD, Kale S et al (2018) Dissecting genomic hotspots underlying seed protein, oil, and sucrose content in an interspecific mapping population of soybean using high-density linkage mapping. Plant Biotechnol J 16:1939–1953. https://doi.org/10.1111/pbi.12929

Patra JK, Das G, Paramithiotis S et al (2016) Kimchi and other widely consumed traditional fermented foods of Korea: a review. Front Microbiol 7:1493

Pelto L, Isolauri E, Lilius EM (1998) Probiotic bacteria down-regulate the milk-induced inflammatory response in milk-hypersensitive subjects but have an immunostimulatory effect in healthy subjects. Clin Exp Allergy 28:1474–1479

Pham AT, Shannon JG, Bilyeu KD (2012) Combinations of mutant FAD2 and FAD3 genes to produce high oleic acid and low linolenic acid soybean oil. Theor Appl Genet. https://doi.org/10.1007/s00122-012-1849-z

Poland JA, Brown PJ, Sorrells ME et al (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PloS One 7:e32253.

Rani A, Kumar V, Verma SK et al (2007) Tocopherol content and profile of soybean: genotypic variability and correlation studies. JAOCS 84:377–383. https://doi.org/10.1007/s11746-007-1040-x

Sayama T, Ono E, Takagi K et al (2012) The Sg-1 glycosyltransferase locus regulates structural diversity of triterpenoid saponins of soybean. Plant Cell 24:2123–2138. https://doi.org/10.1105/tpc.111.095174

Schmutz J, Cannon SB, Schlueter J et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183. https://doi.org/10.1038/nature08670

Severin AJ, Woody JL, Bolon Y-T et al (2010) RNA-Seq atlas of Glycine max: a guide to the soybean transcriptome. BMC Plant Biol 10:160. https://doi.org/10.1186/1471-2229-10-160

Shanmugasundaram S, Yan MR (2010) Vegetable Soybean. In: The soybean: botany, production and uses. CABI, Wallingford, p 427

Shurtleff W, Aoyagi A (2018) The book of miso, 2nd edn. Soyinfo Centre
Sohn SI, Pandian S, Oh YJ, Kang HJ, Cho WS, Cho YS (2021) Metabolic engineering of Isoflavones: an updated overview. Front Plant Sci. https://doi.org/10.3389/fpls.2021.670103
Song Q, Hyten DL, Jia G, et al (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS One. 8(1):e54985. https://doi.org/10.1371/journal.pone.0054985
Song Q, Jenkins J, Jia G et al (2016) Construction of high resolution genetic linkage maps to improve the soybean genome sequence assembly Glyma1.01. BMC Genomics 17:33
Song QJ, Marek LF, Shoemaker RC et al (2004) A new integrated genetic linkage map of the soybean. Theor Appl Genet 109:122–128. https://doi.org/10.1007/s00122-004-1602-3
Song Q, Jia G, Zhu Y et al (2010) Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. Crop Sci 50:1950–1960. https://doi.org/10.2135/cropsci2009.10.0607
Song DH, Chun BH, Lee S et al (2021) Comprehensive metabolite profiling and microbial communities of Doenjang (fermented soy paste) and Ganjang (fermented soy sauce): a comparative study. Foods 10(3):641
Soystats (2022) A reference guide to important soybean facts and figures. American Soybean Association, pp 1–35
Stewart-Brown BB, Song Q, Vaughn JN, Li Z (2019) Genomic selection for yield and seed composition traits within an applied soybean breeding program. G3 Genes Genomes Genet 9:2253–2265. https://doi.org/10.1534/g3.118.200917
Subramanian S, Hu X, Lu G et al (2004) The promoters of two isoflavone synthase genes respond differentially to nodulation and defense signals in transgenic soybean roots. Plant Mol Biol 54:623–639. https://doi.org/10.1023/B:PLAN.0000040814.28507.35
Sugano M, Yamada Y, Yoshida K et al (1988) The hypocholesterolemic action of the undigested fraction of soybean protein in rats. Atherosclerosis 72(2-3):115–122
Sui M, Jing Y, Li H et al (2020) Identification of loci and candidate genes analyses for tocopherol concentration of soybean seed. Front Plant Sci https://doi.org/10.3389/fpls.2020.539460
Sumi H, Hamada H, Nakanishi K, Hiratani H (1990) Enhancement of the fibrinolytic activity in plasma by oral administration of nattokinase. Acta Haematol 84:139–143
Sundaramoorthy J, Park GT, Mukaiyama K et al (2018) Molecular elucidation of a new allelic variation at the Sg-5 gene associated with the absence of group A saponins in wild soybean. PLoS One 13:1–17. https://doi.org/10.1371/journal.pone.0192150
Takagi K, Yano R, Tochigi S et al (2018) Genetic and functional characterization of Sg-4 glycosyltransferase involved in the formation of sugar chain structure at the C-3 position of soybean saponins. Phytochemistry 156:96–105. https://doi.org/10.1016/j.phytochem.2018.09.002
Takahashi K, Banba H, Kikuchi A et al (1994) An induced mutant line lacking the α subunit of β-conglycinin in soybean (*Glycine max* (L.) Merril). Breed Sci 44:65–66
Takahashi J, Furuguchi K, Miyama K et al (1996) Effect of inoculation size on quality of natto. Rpt Food Res Inst Tochigi Prefect 10:17–18
Takahashi M, Uematsu Y, Kashiwaba K et al (2003) Accumulation of high levels of free amino acids in soybean seeds through integration of mutations conferring seed protein deficiency. Planta 217:577–586. https://doi.org/10.1007/s00425-003-1026-3
Takamatsu K (2005) Soy peptides as functional food faterial. In: Soy in health and disease prevention. CRC Press, pp 235–249
Terada A, Yamamoto M, Yoshimura E (1999) Effect of the fermented soybean product "natto" on the composition and metabolic activity of the human fecal flora. Jpn J Food Microbiol 16:221–230
Tettelin H, Masignani V, Cieslewicz MJ et al (2005) Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proc Natl Acad Sci U S A 102:13950–13955. https://doi.org/10.1073/pnas.0506758102
Thomson MJ (2014) High-throughput SNP genotyping to accelerate crop improvement. Plant Breed Biotechnol 2:195–212. https://doi.org/10.9787/pbb.2014.2.3.195

Torkamaneh D, Laroche J, Valliyodan B et al (2021) Soybean (*Glycine max*) Haplotype Map (GmHapMap): a universal resource for soybean translational and functional genomics. Plant Biotechnol J 19:324–334. https://doi.org/10.1111/pbi.13466

Tsukamoto C, Shimada S, Igita K et al (1995) Factors affecting isoflavone content in soybean seeds: changes in isoflavones, saponins, and composition of fatty acids at different temperatures during seed development. J Agric Food Chem 43:1184–1192. https://doi.org/10.1021/jf00053a012

Ujiie A, Yamada T, Fujimoto K et al (2005) Identification of soybean varieties with high α -tocopherol content. Breed Sci 125:123–125

Valliyodan B, Cannon SB, Bayer PE et al (2019) Construction and comparison of three reference-quality genome assemblies for soybean. Plant J 100:1066–1082. https://doi.org/10.1111/tpj.14500

Van Eenennaam AL, Lincoln K, Durrett TP et al (2003) Engineering Vitamin E content: from Arabidopsis mutant to soy oil. Plant Cell 15:3007–3019. https://doi.org/10.1105/tpc.015875

Watanabe D, Adányi N, Takács K et al (2017) Development of soybeans with low P34 allergen protein concentration for reduced allergenicity of soy foods. J Sci Food Agric 97:1010–1017. https://doi.org/10.1002/jsfa.7827

Wilson S, Blaschek K, Gonzalez De Mejia E (2005) Allergenic proteins in soybean: processing and reduction of P34 allergenicity. Nutr Rev 63:47–58. https://doi.org/10.1301/nr.2005.feb.47-58

Wu D, Li D, Zhao X et al (2020) Identification of a candidate gene associated with isoflavone content in soybean seeds using genome-wide association and linkage mapping. Plant J 104: 950–963

Yamamoto T (2005) Soybean components and food for specified health uses (FOSHU). In: Soy in health and disease prevention. CRC Press, pp 279–288

Yang A, Yu X, Zheng A, James AT (2016) Rebalance between 7S and 11S globulins in soybean seeds of differing protein content and 11SA4. Food Chem 210:148–155. https://doi.org/10.1016/j.foodchem.2016.04.095

Yano R, Takagi K, Takada Y et al (2017) Metabolic switching of astringent and beneficial triterpenoid saponins in soybean is achieved by a loss-of-function mutation in cytochrome P450 72A69. Plant J 89:527–539. https://doi.org/10.1111/tpj.13403

Yi J, Derynck MR, Li X et al (2010) A single-repeat MYB transcription factor, *GmMYB176*, regulates CHS8 gene expression and affects isoflavonoid biosynthesis in soybean. Plant J 62: 1019–1034

Yoshikawa M, Tsuruki T (2005) Bioactive peptides derived from soy protein. In: SOY in health and disease prevention. CRC Press, pp 225–233

Nutraceutical Potential of Rapeseed: Breeding and Biotechnological Approaches

Mehak Gupta and Gurpreet Kaur

Contents

Abstract

Rapeseed (*Brassica napus* L.) is a prime oil crop of the world that also provides proteins for the livestock feeding. This crop has achieved remarkable success over the past few decades by development of canola types with low erucic acid

M. Gupta · G. Kaur (✉)
Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, India
e-mail: gurpreetkaur@pau.edu

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_13

and glucosinolate contents. Considerable progress has also been made to enhance the oleic acid content of seed oil. In contrary, little attention has been paid to improve other health-promoting constituents such as phytosterols and tocopherols of the oil. Following oil extraction, rapeseed meal is an excellent source of vitamins, minerals, and high-quality proteins that can help to mitigate human malnutrition. However, poor digestibility, dark color, and bitter taste of rapeseed meal due to anti-nutritional factors like phenolics, phytates, and crude fiber make the meal unacceptable for human consumption. It is crucial to further improve the nutraceutical and commercial value of rapeseed by increasing the content of health-benefitting components while simultaneously minimizing the antinutritive components to the acceptable amounts. To achieve the objective, it is imperative to understand the consolidated genetic architecture of oil and meal quality traits in connection to each other. Recent advances in next-generation sequencing technologies, availability of pan-genome of *B. napus* together with improved bioinformatic and genome editing methodologies would be very useful to reveal the genetic networks and identify high-resolution sequence-based markers for marker-assisted breeding for quality traits in rapeseed.

Keywords

Rapeseed · Nutraceutical · Marker-assisted selection · Erucic acid · Glucosinolates

1 The Crop Rapeseed

Brassica napus L. (AACC, 2n = 38) is one of the six species comprising the well-known U's Triangle of *Brassica*, a genus that belongs to the plant family Brassicaceae. It is an allopolyploid crop species that possesses the full chromosome complements of two diploid species, known to be *B. rapa* and *B. oleracea* (Chalhoub et al. 2014). A few interspecific hybridization events followed by genome duplication are believed to occur in the recent past (<10,000 mya), when two progenitor species (*B. rapa* and *B. oleracea*) were growing in close geographical proximity in Mediterranean area to originate *B. napus* (Friedt and Snowdon 2010). *B. napus* encompasses two subspecies – *B. napus* ssp. *oleifera* (commonly known as rapeseed/oilseed rape) and *B. napus* ssp. *rapifera* (swede and rutabaga) cultivated for oil and fodder purposes, respectively. Of these two subspecies, *B. napus* ssp. *oleifera* carries abundant amount of oil (45–50%) in its mature seeds that has both nutritional and industrial applications depending upon fatty acid composition of oil (Nesi et al. 2008). Also, meal/cake, a by-product left after oil extraction, holds great promise as an excellent protein source (38–40%) for animal nutrition. Presently, rapeseed crop represents the third largest valuable source of vegetable oil (after soybean and oil palm) and the second largest source of protein rich extraction meal (after soybean) in the world (Tang et al. 2021). The crop is cultivated over a large geographical spread of around 33 million ha in Europe, America, China, and India. Different ecotypes of rapeseed, i.e., winter,

semi-winter, and spring types, are raised in diverse ecologies of the world (Song et al. 2020). Winter-type rapeseed with strong vernalization requirement is predominantly grown in Europe, parts of China, and eastern USA, whereas, very early flowering spring forms with no vernalization requirement are cultivated in Canada and Northern Europe. The rapeseed cultivation in Australia and China is largely based on semi-winter types that need mild vernalization (Friedt and Snowdon 2010). Winter cultivars give higher seed yield than spring types owing to advantage of lengthy crop season of over 300 days occupied by winter lines in contrast to only 120–190 days taken by spring cultivars. Historically, rapeseed oil was used widely as lamp oil for European domestic and railway coaches lighting or for some other technical purposes, i.e., to produce soaps, inks, candles, and lubricants. It was of limited value for edible use because of the presence of appreciable amount of two anti-nutritional components – erucic acid (EA) and glucosinolates (GSLs) in the seed, which have detrimental health effects on humans and animals, respectively. The huge success of rapeseed crop for edible oil and animal feed was achieved during just the past four decades after dramatic reduction of these two antinutritive components in the seeds. A trade name "Canola" or double zero/"00" was registered in Canada in 1979 that refers to the edible oil crop carrying low EA (<2%) in the oil and low level of GSLs in the extraction meal (<30 μmol/g of defatted meal). Tower was the first canola rapeseed cultivar, commercially released in Canada in 1974. Thereafter, canola forms largely replaced the traditional rapeseed in all rapeseed-producing areas in Australia, Canada, and Europe and to a very large extent in China (Friedt and Snowdon 2010). The cultivation of canola rapeseed has started in India from 2007 (Chauhan et al. 2010).

2 Nutraceutical Profile of Rapeseed Oil and Meal

Rapeseed oil and meal constitute excellent source of nutraceuticals that are crucial for human nutrition and health as discussed below.

2.1 Rapeseed Oil

In the rapeseed oil, there are triacylglycerols (fatty acids), representing the major component (98%) and non-glyceride fractions (phytosterols, tocopherols, pigments (carotenoids and chlorophylls), and waxes), that constitute the remaining minor fraction (2%) of oil.

2.1.1 Fatty Acids

The fatty acid composition which includes saturated fatty acids (palmitic (PA; C16:0) and stearic acids (SA; C18:0), monounsaturated fatty acid (MUFA) such as oleic acid (OA; C18:1), polyunsaturated fatty acids (PUFA) like linoleic (omega-6; LA; C18:2) and linolenic acids (omega-3; LiA; C18:3), and very long-chain fatty acids (VLCFA) like eicosenoic acid (C20:1) and erucic acid (EA; C22:1) determines the nutritional quality of rapeseed oil as seen in other vegetable oils. The rapeseed oil carries a healthy

lipid profile, fit for human consumption (Russo et al. 2021). The rapeseed oil is low in saturated fatty acids (<7%) and high in MUFA and PUFA, with OA accounting for 61%, LA for 21%, and LiA for 11%, that help in balancing cholesterol level. LA and LiA are two essential fatty acids as they are not synthesized by our body and need to be obtained from the diet. Rapeseed oil contains considerable amount of these two fatty acids with a nearly ideal ratio (5–10:1) to avert cancer, heart, and autoimmune-related diseases. High content of MUFA imparts stability to oil by making it resistant toward oxidation at high temperature and ambient storage. Earlier, rapeseed oil was not acceptable for edible purpose in many countries around the world due to the occurrence of high level of EA (>50%), which is believed harmful to health. EA has slow tendency to oxidize, thus, it accumulates in arterial lining resulting in fibrotic heart lesions. It was found responsible for body weight reduction and abnormal accumulation of fat in laboratory rats (Charlton et al. 1975). Although the negative effects of high EA have never been observed in human beings, it is still suggested to keep the level of EA low (<2%) in rapeseed oil for safe human consumption. Conventionally bred canola rapeseed cultivars lack this nutritionally undesirable long-chain fatty acid (<2% of C22:1), thus making the oil suitable for human consumption.

2.1.2 Phytosterols

Phytosterols in rapeseed occur as free sterols or combined in esters, glucosides, and esterified glucosides. Phytosterols comprise of 45–60% sitosterol, 29–43% campesterol, and 5–13% brassicasterol, with minor fractions of avenasterol (3–7%), stigmasterol (<1%), and cholesterol (<1%). Brassicasterol is a distinct type of phytosterol reported only in Brassicaceae family. Significant amount of phytosterols is present in oil of rapeseed cultivars (0.5–1.1%) (Piironen et al. 2000). Rapeseed oil is reported to contain highest phytosterol content just next to corn oil as revealed by biochemical analysis on ten different oil types (Gordon and Miller 1997). Phytosterols are known to reduce low-density lipoprotein cholesterol and total cholesterol (treatment of hypercholesterolemia) in humans, thus reducing the risk of heart diseases (Cabral and Klein 2017). They are proved to play a vital role as anticarcinogenic, anti-inflammatory, and anti-oxidation agents. They are also known to impart beneficial effects on dementia. Phytosterols are the essential membrane constituents. They protect the skin from UV rays by quenching singlet oxygen. Phytosterols need to be procured from the diet as they cannot be synthesized by our body.

2.1.3 Tocopherols

Rapeseed oil is one of the richest natural sources of tocopherols. The major form of tocopherol in rapeseed oil is γ-tocopherol followed by α- and δ-tocopherols. α-tocopherol is the biologically active form of vitamin E. Vitamin E has potential antioxidant properties and are helpful in avoiding aging-related diseases, Alzheimer's disease, and cardiovascular diseases (Gugliandolo et al. 2017). An adequate intake of vitamin E can also help to reduce chances of neurological disorders, cataracts, and cancer (Schneider 2005). γ-tocopherol is also important as it imparts oxidative stability to rapeseed oil at higher temperatures.

2.2 Rapeseed Meal

Hulls (seed coat) and embryos together constitute the de-oiled rapeseed meal with hull contributing nearly 30% of the total meal weight. Most of the carbohydrates including fiber and phenolic compounds are present in hull portion with tiny fraction of them in embryo part. On the other hand, embryo portion is mainly rich in protein which is less in hull part.

2.2.1 Proteins

The seeds of *B. napus* hold near 45% oil and 25% seed storage proteins. After oil extraction, the meal carries up to 40% of protein. Cruciferins-12S globulins (60%) and napins-2S albumins (20%) are the major proteins account for 80–83% of the total protein content (Swati et al. 2015). Other proteins such as oleosins and lipid transfer proteins constitute the minor proportion of total proteins. Rapeseed protein has a high biological value due to its nearly ideal amino acid composition, which is scarce in cereal proteins, making it an excellent addition to cereal-based diets. Slightly lower levels of essential sulfur containing amino acids (methionine and cystine in napin) have been reported in rapeseed protein as compared to soybean protein. Rapeseed meal protein shows amino acid composition comparable with milk protein casein. The cultivation of modern canola cultivars with low glucosinolate level decreased the amount of the desired napin in seeds (Malabat et al. 2003). The pathways of glucosinolate and amino acid biosynthesis are interconnected and share common enzymes, so manipulation of glucosinolate in canola cultivars could affect the content of napin (Nesi et al. 2008). Peptide mixtures of rapeseed are proved to show antidiabetic, anorexigenic, antiviral, anticancer, antioxidant, and bile acid binding properties (Alashi et al. 2014). Presently, the rapeseed meal is being used solely for livestock feeding. However, it can be served as a potential protein source for poultry and for human nutrition as well. To achieve this, there is need to decrease the content of most limiting anti-nutritional components (GSLs, fiber, sinapates, and phytic acid) responsible for unpleasant bitter taste, dark color, and reduced digestibility of rapeseed meal.

2.2.2 Vitamins and Minerals

Besides proteins, the rapeseed meal is also an abundant source of vitamins (A and C) and minerals. The calcium and phosphorus content of 0.64% and 1.03%, respectively, has been reported. These values are relatively higher than present in soybean meal.

2.2.3 Fiber

Rapeseed meal contains approximately 11–13% dietary fiber. Dietary fibers are of two types, soluble dietary fiber (SDF) and insoluble dietary fiber (IDF). IDF contains large amount of lignin and cellulose whereas SDF is comprised of mainly hemicelluloses. IDF can significantly reduce the digestibility of other essential nutrients, although SDF also have considerable ability to reduce the digestibility. The fiber negatively influences the seed oil and protein content, as the biosynthesis of fiber and oil/protein compete for same photosynthetic assimilates. The yellow-seeded genotypes of rapeseed contain less fiber and phenolics and more oil and protein contents

due to thinner and translucent seed coat resulting in a larger embryo and lower hull fraction. So, seed quality in rapeseed can be improved by breeding for yellow color seed coat.

2.2.4 Glucosinolates (GSLs)

GSLs are diverse glucose and amino acid derived secondary metabolites, specific to order Brassicales only. They are nitrogen and sulfur containing compounds comprising a thioglucose and a sulfonated oxime linked to the chain elongated amino acid. Over 130 diverse GSL types have been discovered from 16 dicot angiosperms, most of which are raised for edible purposes (Blažević et al. 2020). Based upon precursor amino acid, GSLs can be aliphatic (derived mainly from methionine), aromatic (phenylalanine or tyrosine derived), and indolic (tryptophan derived) (Halkier and Gershenzon 2006). However, this classification has little significance. Later, Blažević et al. (2020) gave classification based on three criteria, i.e., amino acid precursor, degradation product, and presence or absence of an aromatic moiety. Using these criteria, 130 GSL types were divided into 9 panels from A to I. Fifteen main GSL types are recorded in rapeseed, achieving content ranging from 60 to 100 μmol/g dry weight in seeds. Aliphatic GSLs are the predominant ones, comprising up to 92% of whole GSL types. The GSL biosynthetic pathway can be divided into three steps: (1) chain elongation, i.e., addition of a methylene group into the chain elongated amino acids, (2) formation of core structure of GSL, i.e., addition of the sulfur group and S-glucosylation, and (3) modification of the core structure (i.e., secondary side chain modifications) – benzoylation, desaturation, hydroxylation, methoxylation, and oxidation depending upon GSL type. Intact GSLs are chemically stable and relatively biologically inactive. However, the enzymatic hydrolysis of GSLs by endogenous myrosinase upon physical injury or tissue damage (cutting, chewing, and mixing) releases biologically active compounds such as isothiocyanates (ITC), oxazolidine-2-thiones, thiocyanates (SCN), nitriles (NI), and epithionitriles (Blažević et al. 2020). Isothiocyanates and its derivatives are known for their therapeutic properties such as biocidal, chemopreventive, antioxidant, and antimutagenic effects (Grundemann and Huber 2018). GSLs and their hydrolysis products also known to protect the plants against bacterial and fungal pathogens and generalist herbivores. Some GSL types such as sulforaphane and indole-3-carbinol possess potential anticarcinogenic properties. However, hydrolysis of other GSLs such as progoitrin and epiprogoitrin gives goitrogenic products. Numerous adverse metabolic effects such as reduced appetite, retarded growth, and liver and kidney damage have been recorded in poultry, fish, and pigs (Kaiser et al. 2021). Studies proving the harmful effects of GSLs in humans are limited. Traditional rapeseed carries high level of GSLs (>50 μmol/g of defatted meal). However, canola forms of rapeseed have a much reduced content of GSLs (<30 micromoles/gm of defatted meal), thus making the meal fit for animal nutrition.

2.2.5 Phenolic Compounds

The main phenolics accumulate in rapeseed are phenolic acids along with soluble and insoluble forms of condensed tannins. The phenolic acids exist in the three forms – free, esterified, and insoluble-bound form. Free phenolic acids comprise about 15% of the total phenolic acid content, whereas nearly 80% of total phenolic acids occur in the

esterified form. Sinapic acid is the prime phenolic acid that constitutes around 90% of the total present free phenolic acids and 70.9–96.7% of esterified phenolic acids (Naczk et al. 1998). Tannins are widely distributed complex polyphenolic compounds showing molecular weights ranging from 500 to 6000 Da. They can be classified into insoluble and soluble forms as per their structure and reactivity response to hydrolytic agents. Majority of the tannins in oilseed rape are proanthocyanidins, also known by the name of condensed tannins. They are formed by polymerization of flavan-3-ols or flavan-3,4-diols. The content of tannins in rapeseed meal varies from 0.2% to 3% of defatted meal (Naczk et al. 1998). Rapeseed meal accumulates more phenolic compounds than any other oilseed (Nowak et al. 1992). Phenolic components contribute to astringency, off-flavor, and dark color of meal. Sinapine, an unpalatable phenolic compound, is reported to impart adverse effects on feed intake and body weight gain with diets containing rapeseed meal (Tangtaweewipat et al. 2004). Egg tainting has been observed when tannin-rich rapeseed meal is fed to laying hen (Shahidi and Naczk 1992). Phenolics are known to affect overall protein digestibility and amino acid availability. However, phenolic compounds are also known to possess antioxidant properties and work as radical scavengers and quenchers. They exhibit protection against heart diseases. Rapeseed extracts with sinapine showed excellent antioxidant activities towards liposomes oxidation and low-density particles. Nowak et al. (1992) have reported phenolic compounds to carry antimicrobial properties, effective against gram-negative bacteria (*E. coli* and *Enterobacter aerogenes*) and gram-positive bacteria (*Bacillus subtilis* and *Streptococcus lactis*).

2.2.6 Phytates

In rapeseed, majority of organic phosphorus (up to 90%) is stored in embryo contained phytic acid. The content of phytic acid ranged from 2% to 4% in seeds, 2–5% in the oil-extracted meal, and 5–7% for the protein concentrates of rapeseed (Lickfett et al. 1999). It is biosynthesized through two distinct metabolic pathways that involve seven enzymes – *MIPS*, *MIK*, *IMP*, *ITPK*, *IPK2*, *2-PGK*, *IPK1*, and *MRP5*, a multidrug resistance-associated protein. Phytic acid is believed to act as an anti-nutritional component but also has anticancer and antioxidant activities. Phytates have a high tendency to bind to metal ions, such as calcium, phosphorus, magnesium, iron, and zinc, owing to their strong chelating properties, which ultimately reduce the bioavailability of essential minerals. Phytates are poorly metabolized by humans and non-ruminant chicken and pigs. Thus, phytases are added externally to rapeseed meal to improve the digestibility of phytates and availability of free phosphorus. Genetically eliminating or reducing the phytic acid content in the rapeseed would be an effective approach to make rapeseed meal nutritious for human consumption.

3 Growing Importance of Rapeseed Nutraceuticals in Face of Chronic Diseases and Malnutrition

Inadequate intake of vitamins and minerals, often known as micronutrient malnutrition or hidden hunger, has a negative impact on human health. Even mild level of micronutrient deficiency may promote disease development and affect cognitive

ability. Hidden hunger affects around two billion people worldwide, with the African continent and South Asia having the highest frequency. Micronutrient malnutrition often leads to stunted growth in children (starting from fetal development to 4 years of age), as already observed in developing countries (Branca and Ferrari 2002). Chronic conditions/diseases arising as a result of malnutrition can be effectively treated with the healthy plant-based diets and via biofortification of the important staple crops. Biofortification is the method of breeding food crops that are naturally enriched with micronutrients/macronutrients in the adequate quantity in their edible parts (in a bioavailable form) to mitigate the hidden hunger. Now, the focus of agricultural sector has been inclined towards developing nutrient supplement food crops rather than just increasing the supply of food crops. Biofortification approach targets on enhancing the nutritional value of the food crops that people already consume rather than attempting to add some health-promoting components to the diet. This approach is certainly the most sustainable and cost-effective approach as it is a one-time investment and gives long-term results to alleviate nutrient deficiency. The rapeseed crop has gained substantial importance over the years owing to the health-promoting effects of various bioactive compounds present in it as already mentioned above. Great efforts have been done in the past to improve the composition of rapeseed oil and meal to meet the dietary requirements of the consumers. Levels of anti-nutritional compounds such as EA and GSLs have been reduced to acceptable amounts. However, further interventions are required to convert this crop species into a high-quality food-grade crop in which desirable components should be high enough to support human and animal health but at the same time should not be too high that they become toxic. The prime objectives to enhance the nutraceutical value of rapeseed are to improve the content and composition of the seed oil and meal for food and feed purposes.

4 Attempts to Enhance Seed Oil Content in Rapeseed

Increase in oil content has been one of the most important goals of rapeseed breeding. Variation for seed oil content has been reported in *B. napus* germplasm. It varies from 26% to 50% (Xiao et al. 2019). The trait is under the control of complex regulatory mechanisms that is still not completely understood. Genetic studies have revealed the polygenic inheritance of the trait, which is frequently influenced by the environment.

4.1 Mapping for Oil Content

Numerous QTLs (quantitative trait loci) ranging from 3 to 27, each with small effect of $<10\%$, have been detected to control the variation for oil content in *B. napus* in many biparental QTL mapping studies (Zhao et al. 2012; Jiang et al. 2014). These have been visualized on all 19 chromosomes of *B. napus*. However, many QTLs,

i.e., 67, including a few QTLs with more than 20% phenotypic variation explained, were identified in a doubled haploid (DH) population generated from the cross between KenC-8 and N53-2 varieties of *B. napus* (Chao et al. 2017). The study also found some common QTLs that affected both oil and protein contents but in an opposite manner. Four QTLs on chromosomes A08, A09, C03, and C06, explaining >10% effects for oil content consistently over more than four trails, were further screened to detect candidate genes controlling oil content by integrating genome resequencing and transcriptomic analysis of parental genotypes (Yan et al. 2022). More recently, genome-wide association studies (GWAS), which exploit historical recombination events and provide high mapping resolution, have been attempted to reveal the genetic basis and identify candidate genes for seed lipid content in *B. napus.* A patatin-like lipase (*PTL*) gene on C07 was found to be associated with oil content using GWAS in a collection of 290 rapeseed germplasm accessions (Wang et al. 2021). Another major QTL (*qA07.SOC*) on A07 has been detected for seed oil content based on the GWAS and RIL population (derived from *B. napus* and *Sinapis alba*) based mapping (Zhao et al. 2022). Tang et al. (2021) also provided a deeper understanding of oil accumulation in rapeseed by associating genome variants and transcriptomes (of developing seeds) of a large natural *B. napus* population (505 inbred lines) with multiyear, multilocation seed oil content data. Previously reported 27 QTLs and some novel QTLs for seed oil content were detected in the study. QTL regions under breeding selection were also underpinned based on selection intensity of identified QTLs. A pair of homologous genes, *BnPMT6s* encoding a S-adenosyl-L-methionine-dependent methyltransferase, was demonstrated to negatively affect the oil content.

4.2 Characterization of Genes Involved in Oil Biosynthesis

A few genes regulating oil content have been cloned in *B. napus* (Liu et al. 2019). *ORF188*, a novel chimeric mitochondrial gene cloned using comparative genomics and transcriptome analysis, increased oil content by 8% in *B. napus* (Liu et al. 2019). Ding et al. (2019) used RNA interference (RNAi) technology to elucidate *BnLACS2* to be the key element for seed oil production. *BnLACS2* is a long-chain acyl-CoA synthetase, whose overexpression increased the oil content, whereas its silencing lowered oil content in rapeseed plants. Transgenic oilseed rape lines showed a near 10% increase in oil content by overexpression of a lysophosphatidate acyltransferase (*LPAT*) gene from yeast (*Saccharomyces cerevisiae*) (Zou et al. 1997; Taylor et al. 2002). About 40% surge in oil content has been recorded by an increase in Gly3P content upon overexpression of a yeast *Gly3P* dehydrogenase gene in rapeseed (Vigeolas et al. 2007). Seed-specific overexpression of the *BnLEC1* and *BnL1L* genes at an appropriate level under the control of the two truncated canola storage protein 2S-1 promoters, also known as the *napA* promoters, significantly promoted the seed oil content in rapeseed with no detectable negative effects on other major agronomic traits, including protein content (Tan et al. 2011). The CRISPR-CAS9 has emerged as a powerful technology to study gene function and produce desired traits

by efficient mutagenesis of targeted genes (Zhang et al. 2021). The technology has been used recently for functional characterization of genes involved in oil biosynthesis. Knocking out many *BnSFAR4* and *BnSFAR5* genes at the same time improved seed oil content of *B. napus* without affecting seed germination or vigor (Karunarathna et al. 2020). Silencing of the *BnLPAT2* and *BnLPAT5* genes caused decreased oil content and enlarged oil bodies (Zhang et al. 2019).

5 Improvement in Fatty Acid Composition of Rapeseed Oil

5.1 Reduction in Erucic Acid (EA)

EA is synthesized from OA by elongation process catalyzed by a rate-limiting enzyme *β*-ketoacyl-CoA synthase (KCS) encoded by the gene *FAE1*. Mutants for FAE1 led to decreased content of longer chain fatty acids (Lassner et al.1996). A mutant identified from German spring rapeseed cultivar 'Liho' is regarded as the only donor source for low EA to all canola quality rapeseed cultivated today. 'Oro' and 'Zephyr' were the first low EA rapeseed cultivars, followed by many, that were developed through conventional breeding efforts and released first in Canada. Content was found to regulate by two loci with additive effect as F_1 seeds from cross between low and high EA genotypes depicted intermediate EA content between corresponding parents and segregation ratios in F_2 and F_3 seeds were in accordance to expected ratios under control of two genes (Fourmann et al. 1998).

5.1.1 Mapping for EA Content

Various molecular markers have been used for mapping the EA content. Two additive loci (E^A and E^C on A and C genomes of *B. napus,* respectively) were identified that on collective basis explained 90% of whole EA variation in *B. napus*, while contributing unequally to the final EA content (Jourdren et al.1996). The two genes *BnFAE1.1* on A08 and *BnFAE1.2* on C03*,* encoding *β*-ketoacyl-CoA synthases, were cloned and revealed to co-segregate with E^A and E^C, respectively (Fourmann et al. 1998). GWAS studies have also been carried out to identify loci for EA content of rapeseed oil. Candidate genes *BnaA08g11130D* and *BnaC03g65980D*, orthologous to Arabidopsis *FAE1*, and *BnaA08g11140D* and B*naC03g66040D*, orthologous to *KCS17*, were identified through GWAS (Zhu et al. 2019).

5.1.2 Characterization of *FAE* Genes

There are six gene copies encoding *FAE1* proteins in *B. napus* (Wu et al. 2008; Cao et al. 2010). Of these, *BnaA8.FAE1* and *BnaC3.FAE1* are the major ones that control erucic acid synthesis in seeds. Comparison of *FAE1* sequences between HEA (high erucic acid) and LEA (low erucic acid) genotypes of *B. napus* indicated nucleotide variations. Functional activity of *FAE1* was lost *via* deletion of two to four bases in C genome homolog (Fourmann et al. 1998; Wu et al. 2008) or substituting C/T in A genome homolog (Han et al. 2001). Single SNP substitution causing amino acid change from serine to phenylalanine led to the

development of LEA *B. napus* cv. Oro (Katavic et al. 2002). Nineteen mutants for the *BnFAE1* fragment were detected through the targeting induced local lesions in genomes (TILLING) approach during the screening of 1344 M_2 plants. Of these, three M_3 plants showed reduced EA (Wang et al. 2010). The low level of EA content in *B. napus* was achieved by inhibiting the gene expression of *FAE1* genes through RNAi (Peng et al. 2010; Shi et al. 2017) and CRISPR/Cas9 technologies (Liu et al. 2022a).

5.2 Progress to Increase Oleic acid (OA) or/and Reduce Linolenic Acid (LiA) in Rapeseed: HOLL (High Oleic and Low Linolenic) Varieties

Increased levels of MUFA (OA) and reduced content of PUFA (LiA) in rapeseed oil are vital to impart thermal stability to the oil so that it can be utilized for food applications that need high cooking and frying temperatures and for salad dressings. Presently, canola quality rapeseed cultivars contain approximately 55–65% OA. There is a necessity to further increase the OA content in rapeseed oil. Thus, breeding rapeseed cultivars with high oleic acid (HOA) and low linolenic acid (LLiA) content is an important objective. The objective can be achieved by developing high OA germplasm combined with improved understanding of genetic control of trait. In the fatty acid biosynthesis pathway, activity of fatty acid desaturases (*FADs*) regulates the amount of OA, LA, and LiA. The OA is desaturated to LA by inserting double bond at delta 12 position mediated by an enzyme oleoyl-PC D12-desaturase encoded by gene *FAD2* (Hu et al. 2006; Peng et al. 2010; Yang et al. 2012), whereas LA is further desaturated to LiA by inserting double bond at the delta 15 position mediated by another enzyme, delta-15 desaturase encoded by gene *FAD3* (Yang et al. 2012; Lee et al. 2016). HOA and LLiA rapeseed mutants were identified in the physically and chemically mutagenized populations of rapeseed (Bai et al. 2019; Fu et al. 2021). Mutants harboring >80% OA were demonstrated to display reduced agronomic potential (lower seedling vigor, delayed flowering, and reduced plant height) and oil content (Bai et al. 2019). Depending on mutants, one or two major loci for *FAD2* controlling OA (Falentin et al. 2007), while two loci for *FAD3* genes for LiA were revealed (Barret et al. 1999). The first LLiA *B. napus* cultivar, Stellar, was identified in 1987 (Scarth et al. 1988) followed by Scarth with very LLiA (Scarth et al. 1995).

5.2.1 Mapping for OA and LiA Content

Both linkage and association studies have been used to map QTLs for OA in *B. napus* (Hu et al. 2006; Yang et al. 2012; Wen et al. 2015; Qu et al. 2017; Bao et al. 2018; Fu et al. 2021). The QTLs detected through linkage analysis were mapped on chromosomes A01, A05, C01, and C05 of *B. napus*. GWAS for fatty acid composition in a set of 520 genetically diverse rapeseed accessions identified a total of 62 MTAs distributed over 18 chromosomes which explained 2.31–14.48% of the phenotypic variance for seven fatty acids: PA, SA, OA, LA, LiA, EiA, and EA (Qu et al. 2017). However, only five genomic regions present on chromosomes, A02, A08, A09, C01, and C03,

were common in the two years. Six haplotype regions on chromosomes A02, A07, A08, C01, C02, and C03 were identified as significantly associated with OA through GWAS of haplotype blocks (Yao et al. 2019). Further, whole-genome sequencing of 50 rapeseed accessions revealed three genes: *BnmtACP2, BnABCI13*, and *BnECI1* in the haplotype region of A02 and two genes, *BnFAD8* and *BnSDP1* in the C02 responsible for variation in OA content. Association studies in a diversity set comprising 435 rapeseed accessions identified 149 quantitative trait nucleotides (QTNs) for fatty acid content and composition, of which 34 QTNs were overlapped with previously reported loci (Guan et al. 2019). Candidate genes *BnaA08g08280D* and *BnaC03g60080D* for OA indicated genome sequences differences between HOA and LOA genotypes. A novel QTL for *FAD2* explaining 6.2–11.7% of phenotypic variation over three crop seasons was revealed on A09 through GWAS analysis using 375 low EA *B. napus* genotypes. Transfer of this region together with previously mapped QTLs could facilitate increase in OA to 80%. Additionally, fine mapping of identified QTL unraveled three promising candidate genes that could also be validated via gene expression studies. Closely linked marker, *BnA129,* was also designed to enable marker-aided selection for OA (Zhao et al. 2019). Candidate gene-based association study in 324 rapeseed accessions identified three candidate genes present on chromosomes A07, A08, and C03 to be associated with oleic acid (Zafar et al. 2020). In the abovementioned studies, two major QTLs located on chromosomes A05 (Hu et al. 2006; Yang et al. 2012) and C01 (Hu et al. 2006) were found to harbor homologs to *AtFAD2* gene of Arabidopsis, that catalyze OA into LA acid and hence plays crucial role in controlling OA content in seed oil.

5.2.2 Characterization of *FAD* Genes

Four *AtFAD2* orthologs (the major effect *BnFAD2.C01* and *BnFAD2.A05,* plus their homeologous copies *BnFAD2.A01* and *BnFAD2.C05*) have been identified in allotetraploid *B. napus* (Yang et al. 2012; Gacek et al. 2017). Three copies of *BnFAD2* (*BnFAD2.C01*, *BnFAD2.A05*, and *BnFAD2.C05*) are functional while one copy *BnFAD2.A01* seems to be a pseudogene. Mutations in the *BnFAD2* copies have been revealed to influence OA quantity. A single nucleotide polymorphism (SNP) in the coding region of copy *BnFAD2.C01* caused surge in OA content up to 77%, whereas for *BnFAD2.A05* copy, a single-nucleotide substitution or a 4-bp insertion (Hu et al. 2006; Yang et al. 2012) in the coding region caused increase in OA quantity up to 75%. Fu et al. 2021 also confirmed the role of *BnFAD2* copies to control OA by revealing SNPs in *BnFAD2.A05* and *BnFAD2.C05* copies. Loss of function of *BnFAD2* copies in *B. napus* via RNAi (Peng et al. 2010), gene knockouts (Wells et al. 2014), and CRISPR/Cas9-mediated genome editing (Okuzaki et al. 2018; Huang et al. 2020; Liu et al. 2022b) again confirmed the significance of *FAD2* genes in increasing OA amount up to 84–85%. CAPS markers (Falentin et al. 2007), Illumina qPCR-based assay (Hu et al. 2006), SNP markers (Yang et al. 2012), and KASP (Fu et al. 2021) markers have been developed for marker-assisted selection of desirable *FAD2* alleles for high OA. Gene sequence comparison of *FAD3*copies also indicated single nucleotide mutations. For direct selection of *FAD3* allele, restriction site generating–polymerase chain reaction (RG-PCR) (Barret et al. 1999) and SNaPshot (Spasibionek et al. 2020)

have been developed. These markers have been successfully employed for selection of high OA and LLiA content in winter rapeseed.

5.3 Increase in Eicosapentaenoic Acid (EPA) and Docosahexaenoic Fatty Acid Levels

Long-chain unsaturated fatty acids (i.e., EPA and DHA) rich oil is in high demand because of its numerous health benefits. These fatty acids were once acquired exclusively from fish oil. Attempts were made to obtain EPA and DHA from rapeseed oil through transgenic approach. Using six different gene constructs from *Thraustochytrium* sp., *Pythium irregular*, and *Calendula officinalis*, levels of arachidonic acid (AA) and EPA were increased up to 25% and 15% of total seed fatty acids in rapeseed, respectively (Wu et al. 2005). Similarly, transgenic *B. napus* expressing microalgal PUFA synthases (*OrfA*, *OrfB*, and hybrid *OrfC*) from *Schizochytrium* sp. ATCC 20888 showed 2.87–3.43% average DHA content in T2 seeds of inbred lines of selected transformation events (Walsh et al. 2016). The total DHA and EPA content was around 4.4% in the field-produced canola oil (Walsh et al. 2016). In the latest transgenic DHA canola variety, seven fatty acid biosynthesis genes from yeast and microalgae were used as single fragment of around 19 Kb to transform *B. napus*, which increased the DHA levels from 9% to 11% like those obtained from fish (Petrie et al. 2020).

6 Minor Oil Components

In addition to total oil content and fatty acids composition, some minor oil components, such as phytosterols (Amar et al. 2008) and tocopherols (Fritsche et al. 2012; Gugliandolo et al. 2017), have also recently drawn the interest among rapeseed researchers to study and improve their content and composition owing to numerous health-benefitting potentials associated with them. The studies executed on phytosterols and tocopherols are discussed below.

6.1 Phytosterols

Abidi et al. (1999) recorded phytosterol content varying from 7659 to 14,023 mg/kg oil in 12 different spring canola varieties. In another study, a twofold variation was observed in the phytosterol content, varying from 4475 to 9380 mg/kg oil among the lines of three winter-type rapeseed DH populations (Amar 2007). Canola cultivars have been observed to contain higher (3565–4800 mg/kg) phytosterol content than traditional non-canola (2079–4329 mg/kg) cultivars (Amar et al. 2009). In a winter rapeseed DH population segregating for EA, QTL mapping results depicted colocalization of two of the three QTL identified for total phytosterol content and two of the four QTL detected for sinapate ester content with two EA genes on A08

and C03 (Amar et al. 2008). A close negative association of phytosterol and sinapate ester content with EA content was found. The study suggested the pleiotropic effect of the two EA genes on phytosterol and total sinapate ester content in seeds of rapeseed. This was supported by the fact that common cytoplasmic acetyl-CoA precursor is required for the synthesis of both EA and phytosterols. To further understand the genetics of phytosterols and their association with other important seed quality traits (oil content, fatty acids content, protein content of the defatted meal, and seed weight), a DH population constructed from the two-canola quality winter rapeseed cultivars, Sansibar and Oase, was used for QTL mapping (Teh and Mollers 2016). The parental genotypes were differed with respect to both oil and phytosterol content. Variable number of QTLs ranging from one to six were identified for the evaluated traits. QTLs governing large effects ($R^2 \geq 25\%$) were all identified on the A genome. QTLs identified for one trait often found to colocalize with QTLs for other traits.

6.2 Tocopherols

Rapeseed germplasm has been investigated for variation in tocopherol content and composition in some instances. Significant variation in alpha, gamma, and total tocopherol content that varied from 63 to 157 ppm, 114–211 ppm, and 182–367 ppm, respectively, has been revealed in a diverse panel comprising 87 winter rapeseed genotypes in a study by Goffman and Becker (2002). Alpha−/gamma tocopherol content ratio found to be varied from 0.36 to 1.23. However, groups with different seed quality types (double zero, single zero, and conventional) showed no significant differences for tocopherol content and composition. The study showed lack of correlation between alpha and gamma tocopherol contents. However, a positive correlation of 0.34 was observed between gamma-tocopherol content and oil content. In another study, a broader range of variation (197.5–460.1 ppm) was reported for total tocopherol in a worldwide germplasm collection of 299 rapeseed accessions (Fritsche et al. 2012). A candidate gene-based association mapping study performed on 96 accessions of this wide set identified between 26 and 12 SNPs within tocopherol biosynthesis genes (*BnaX.VTE3.a* and *BnaA.PDS1.c*) that explained 16.93% of total genetic variance for tocopherol composition and up to 10.48% for total tocopherol content. Gene-based markers were also designed and validated on the remaining accessions of the set to facilitate the selection of rapeseed genotypes with enhanced tocopherol qualities. Endrigkeit et al. (2009) cloned the first gene *BnaA.VTE4,* involved in tocopherol biosynthesis pathway in rapeseed using sequence information of Arabidopsis *VTE4* orthologs. The function of gene was also verified by an *A. thaliana* transgenic approach leading to 50-fold increase of α-tocopherol in seeds of *BnaA.VTE4.a1* overexpressing plants. A marker assay was developed for the gene and mapped to the position of two QTLs on chromosome A02 of Tapidor X Ningyou7 genetic map to facilitate marker-assisted selection for increased tocopherol content. Wang et al. in 2012 used Tapidor X Ningyou7 derived DHs and its reconstructed F_2 population and a panel of 142 rapeseed accessions to

decipher the genetic control of seed tocopherol amount and composition. They found 33 unique QTLs contributing towards phenotypic variation in tocopherol content and composition in biparental populations. Furthermore, they shortlisted 7 QTLs that colocalized with potential candidate sequences in association with tocopherol biosynthesis based on in silico and linkage mapping studies. GWAS performed on diverse panel of 142 rapeseed accessions revealed 61 loci linked with tocopherol content and composition out of which 11 were located within the regions of QTLs detected in biparental populations. The relative abundance of individual tocochromanol species in the seed oil has been modified by overexpression of chimeric TC gene constructs in developing seeds of transgenic rapeseed plants (Kumar et al. 2005). Co-overexpression of *PDS1*and *VTE2* genes have also been attempted in *B. napus* to augment total tocopherol content (Raclaru et al. 2006).

7 Progress to Improve Rapeseed Meal Value

At present, the use of rapeseed meal is restricted to feed of ruminants only, but it can be used as a high-quality protein source for poultry and hog nutrition and even for human consumption by reducing the content of most limiting anti-nutritional factors such as GSLs, fiber, phytic acid, sinapates, and non-starch polysaccharides (NSP) (Wanasundara et al. 2016). Therefore, the present objective in rapeseed meal improvement is to decrease these undesirable components in the seed. Concurrently, there is a great interest to elevate the total seed protein content and manipulating the protein composition by altering napins and cruciferins ratio.

7.1 Reduction of GSLs

Rapeseed meals with high GSLs are not considered good for animal health. So, rapeseed breeding programs are always aimed at reducing GSL content in the seeds. Breeding for low GSL content was initiated after the discovery of Bronowski, a Polish cultivar of *B. napus* (Finlayson et al. 1973). The low GSL content of Bronowski was successfully incorporated into elite zero EA germplasm of summer and winter rapeseed through conventional cross-breeding and pedigree selection (Kondra and Stefansson 1970; Wittkop et al. 2009). The reduction in GSL content occurred due to decrease in aliphatic type of GSLs (Kondra and Stefansson 1970). Today all cultivated rapeseed depicts low EA (<2%) and GSL (<15 μmoles/g of defatted meal) levels in the oil and meal, respectively. The lower level of GSLs in seeds has also raised the concern regarding increased susceptibility of cultivars to pests, pathogens, and birds owing to simultaneous decrease of GSL content in leaves. Thus, it would be desirable to improve the protective effects of *B. napus* by genetic manipulation of the leaf GSL profile without affecting seed quality. A few investigations have suggested the lack of correlation between GSLs of seeds and leaves and proposed the possibility of tissue specific manipulation of GSL profiles (Fieldsen and Milford 1994). Although GSLs can be traced in all parts of the plant,

their concentration varies throughout the plant body, i.e., unequal amount is present in different parts of plant that may also differ with respect to plant development stages and in response to various biotic and abiotic factors (Wittstock and Burow 2010). For instance, the GSL content was found higher in seeds as compared to leaves in *B. napus*. GSL content of leaves varied between 0.6 and 6.9, whereas in seed, it varied from 10.8 to 57.9 μmol/g of dry material (Baaij et al. 2018). Moreover, profile of GSL also varies in different plant tissues. Aliphatic GSLs predominates both in leaves as well as in seeds, whereas indole GSLs are found more abundantly in leaves than in seeds (Velasco et al. 2008). The differences of GSL content and profile may be attributed to diverse functions of different parts of plants. Total GSL content show quantitative inheritance. Profiling of GSLs in *B. napus* indicated three main types: gluconapin, glucobrassicanapin, and progoitrin. The content of these GSLs was found to be determined by the genotype of the mother parent. Three, four, and five loci were reported to control gluconapin, glucobrassicanapin, and progoitrin, respectively. Higher content of gluconapin and glucobrassicanapin showed partial and overdominance over low content of GSL, whereas partial dominance was indicated for high content of progoitrin over absence of progoitrin (Kondra and Stefansson 1970).

7.1.1 Mapping for Glucosinolate Content and Composition

To reveal the genetic control of GSL synthesis and accumulation in vegetative tissues or/and seeds of rapeseed, several studies have been carried out using biparental linkage mapping and GWAS approaches. Available reports indicate that four major genomic regions present on chromosomes A02, C02, A09, and C09 to be the most likely regions to control seed GSL content variation in *B. napus*. Two loci have been revealed to control side chain elongation and hydroxylation in leaves or seeds of populations derived from natural and resynthesized *B. napus* genotypes (Magrath et al. 1994; Parkin et al. 1994). Uzunova et al. (1995) identified four QTLs on rapeseed linkage groups 2, 9, 16, and 18, which together accounted for 61.7% of observed phenotypic variation for seed GSLs in a DH population. Similarly, four QTLs were revealed by Howell et al. (2003) in linkage groups 2, 7, 9, and 19 that together explained 76% variation for seed GSL content in two backcross populations. A total of 39 QTLs governing variation for seed GSL content were detected, 28 of those were mainly found on 4 chromosomes (A02, A09, C02, and C09), that explained 10.2–35.3% of phenotypic variation on individual basis in the TNDH population (Feng et al. 2012). A major effect QTL, *qGSL-C2*, explaining 30.88–72.87% of the phenotypic variation was identified on chromosome C02 in a DH mapping population developed from cross of two *B. napus* accessions that differed in their seed GSL contents (Liu et al. 2020a). The gene, *BnaC2.MYB28/HAG1*, orthologous to the *AtMYB28,* responsible for aliphatic GSL synthesis, was also suggested as promising candidate gene for the detected QTL.

Three loci on chromosomes A09, C02, and C09 were found to be in association with seed GSL content by Harper et al. (2012) in a small set of 53 *B. napus* lines using associative transcriptomics. They also revealed that deletions of orthologs of the gene *AtMYB28* on chromosomes A09 and C02 led to low seed GSL content

(Harper et al. 2012). Two *MYB28* copies on chromosomes A09 and C02 were identified as key regulators for aliphatic GSL variation in rapeseed leaves though associative transcriptomic analysis on 288 accessions of *B. napus* (Kittipol et al. 2019). Additionally, *Bna.HAG3/MYB29.A3*, was revealed to control root aromatic GSL variation. A candidate gene *Bna.MAM3.A3* was also proposed to have a role in phenylalanine chain elongation for aromatic GSL biosynthesis based on root expression data. Lu et al. (2014) inferred 26 candidate genes including orthologs of gene *AtMYB28*, i.e., *BnaA.GTR2a* and *BnaC.HAG3b*, for low seed GSL content in a larger set of 101 *B. napus* lines. Further, Li et al. (2014) also suggested that different copies of *AtMYB28* on chromosomes A09, C02, C07, and C09 responsible for the seed GSL content using a Brassica 60 K SNP array on a set of 472 rapeseed genotypes. Wang et al. (2018) identified 49 loci and 27 candidate genes for association with seed GSL content using GWAS in a diversity set. Further, 5 common and 11 tissue-specific loci were associated to total leaf and seed GSL content using GWAS analysis in 366 accessions of *B. napus* (Liu et al. 2020b). A candidate gene *BnaA03g40190D (BnaA3.MYB28)* responsible for high leaf and low seed GSL content was also validated by gene sequence polymorphism and expression assays.

7.1.2 Characterization of GSL genes

Specific gene regulatory mechanisms combined with environmental influences control GSL biosynthesis, transport, and accumulation, resulting in varying content and distinct GSL profile in different tissues of Brassica species (Mitreiter and Gigolashvili 2021). The glucosinolate biosynthetic pathway has been well characterized in the model species *A. thaliana.* Almost all genes involved in three biosynthesis stages, i.e., side chain elongation of amino acids, core structure formation, and modification of secondary side chain, have been revealed (Sonderby et al. 2010). An intricate network of TFs together with biotic and abiotic stimuli and hormonal and epigenetic factors governs the spatiotemporal synthesis of GSLs (Hirai et al. 2007). The three types of GSLs (aliphatic, indolic, and aromatic) are independently regulated by different group of genes (Sonderby et al. 2010). Three TFs, *MYB28, MYB76*, and *MYB29*, belonging to the *R2R3-MYB* family, regulate biosynthesis of aliphatic GSLs (Gigolashvili et al. 2007, 2008a, b). They are also denoted as *HIGH ALIPHATIC GLUCOSINOLATE (HAG) 1, 2,* and *3,* respectively. Of these, TF*MYB28 (HAG1)* is the prime transcriptional regulator that upregulate most of the genes involved in aliphatic GSLs biosynthesis, including side chain elongation (*MAM1* and *MAM3*) and core structure development (*CYP79F1*, *CYP79F2*, *CYP83A1, ST5b*, and *ST5c*) (Gigolashvili et al. 2008a, b; Baskar and Park 2015). Owing to closest evolutionary relationship of *A. thaliana* with Brassica crops, many *MYB28*orthologues have been identified in Brassica species, including four in *B. rapa* (Kim et al. 2013), three in *B. oleracea* (Augustine et al. 2013), five in *B. juncea* (Yin et al. 2017), and six in *B. napus* (Long et al. 2016). The DNA sequences of *MYB28* members in Brassica crops are found to be highly conserved. The coding region varies from 1350 to 1630 bp in length. The inferred *BnMYB28* protein sequences of *B. napus* share percent identity ranging from 59% to 78% with *AtMYB28* and from 59% to 98% with each other (Long et al. 2016). Many genes explaining variation for GSL content in *B. napus* have been identified via

QTL mapping, genome-wide association, and associative transcriptomic studies as already described above. These studies indicated the significant role of *MYB28* homologs in aliphatic glycoside biosynthesis. Aliphatic as well as total GSL content were also reduced by silencing of *MAM* gene family through RNAi (Liu et al. 2011). Independent knockout mutants of genes *MYB28* and *CYP79F1* have been verified to carry decreased aliphatic GSL content by 55.3% and 32.4%, respectively, compared to control plants in *B. napus* (Jhingan et al. 2023). Genes encoding TFs *MYB34/ATR1 (altered tryptophan regulation 1)*, *MYB51 (HIG1)*, *MYB122 (HIG2)*, and cytochrome P450 enzymes, *CYP79B2, CYP79B3*, and *CYP83B1*, are associated with biosynthesis of indolic GSLs (Celenza et al. 2005; Gigolashvili et al. 2007; Frerigmann and Gigolashvili 2014). Overexpression of allele, *atr1D,* altered the expression of *CYP79B2, CYP79B3*, and *CYP83B1*, which led to increased levels of indolic GSLs, whereas, loss of function mutants resulted in decreased levels of indolic GSLs (Celenza et al. 2005). Redirection of tryptophan into tyramine transformed rapeseed with *TDC* (*tryptophan decarboxylase*) gene resulted in lower levels of indolic GSLs (Chavadej et al. 1994). Limited information is available for synthesis of aromatic type of GSLs. It is obvious that GSLs biosynthesis pathway in Brassica crops is extremely complicated than *A. thaliana* due to presence of polyploidization and multiple genomic rearrangements events. Spatial-temporal transcriptional regulation of different *MYB28* homologues has been commonly observed in Brassica, suggesting the occurrence of functional divergence of *MYB28* copies after genome polyploidization (Kim et al. 2013). Genetically, removal of GSLs in seeds while maintaining them in leaves could be achieved in *A. thaliana* through manipulation of two transporters, *GTR1* and *GTR2*, that control GSL accumulation in seeds (Nour-Eldin et al. 2012). GSLs are primarily synthesized in leaves and silique walls (source tissues) and then translocated to embryos via phloem by two transporters, *GTR1* and *GTR2* (Nour-Eldin and Halkier 2009). The *gtr1 gtr2* double mutant of *A. thaliana* failed to accrue GSLs in its seeds; however, it over-accumulated GSLs higher than tenfold in the source tissues, leaves, and silique walls (Nour-Eldin et al. 2012). In the similar way, low level of GSLs has been achieved through gene editing of transporter genes, *GTR 2* gene in *B. napus* (He et al. 2022). Overexpression of STM gene caused reduction in seed GSL levels (Elhiti et al. 2012). Modified GSL content was also achieved by engineering leaf cotyledon 1 gene (Elahi et al. 2016).

7.2 Reduction of Other Antinutritive Compounds

Breeding of rapeseed cultivars with reduced phenolic compounds (sinapates, tannins, and proanthocyanidins), fiber, and phytates in seeds is important as these components limit the usage of rapeseed meal as a quality grade protein source for human and animal consumption. Interestingly, seed coat color is found to be associated with some of these antinutritive seed constituents in Brassica species. Proanthocyanidins and tannins are the prime components responsible for seed coat pigmentation. Black/brown-seeded genotypes have more proanthocyanidins, tannins, and fiber deposited in their seed coats. In contrast,

yellow-seeded lines with thinner seed coats display improved quality of the seed (higher proportions of oil (also more transparent) and protein) and derived meal (lower quantities of tannins, proanthocyanidins, and fiber). Much attempts have been carried out for the development of yellow-seeded genotypes in oilseed rape. There is no naturally occurring germplasm with yellow seed color in *B. napus.* Most of the yellow-seeded *B. napus* has been derived from interspecific/intergeneric crosses of *B. napus* with related species *B. rapa, B. juncea, B. carinata, B. oleracea* spp. *Alboglabra*, and wild crucifers (*Sinapis alba* and *Descurainia sophia*) (Wen et al. 2012). Although seed coat color/seed pigmentation is a morphological marker, it is difficult to select owing to its low heritability, polygenic inheritance, and maternal and environmental effects (temperature and light) controlling the trait (Stein et al. 2013). So, molecular markers linked to seed coat color were also established in rapeseed (Rahman et al. 2010). Transparent testa mutants (TT), including early (EBGs) and late (LBGs) biosynthesis genes, are known to mainly regulate variation in seed coat color (Appelhagen et al. 2014). The EBGs include *TT4, 5, 6*, and *7,* and LBGs include *TT3, 12, 18, 19, BAN*, and *AHA10.* A few seed coat color genes *TT1TT2, TT10, F3'H, PAL, BAN*, and *TTG1* have also been cloned and revealed to be involved in flavonoid biosynthesis pathways in *B. napus* (Zhang et al. 2013; Lian et al. 2016). Concurrently, genetic mechanisms of seed fiber (Behnke et al. 2018; Miao et al. 2019; Gacek et al. 2021) and condensed tannins (Lipsa et al. 2012) have also been studied together with seed color, oil, protein, and glucosinolate content in order to improve oil content and meal value of rapeseed. QTLs of one trait were repeatedly found to be colocalized with other traits because of interlinked biochemical pathways involved in the synthesis of these components from common substrates/precursors. A lignin biosynthesis gene, *BnCCR1*, on A09 was reported to affect both seed color and lignin content (Liu et al. 2012). More recently, a QTL cqSC-A09 has been detected which simultaneously explained variation for seed color, fiber, and oil content with a large effect (Chao et al. 2022). Most of the studies revealed one major locus on chromosome A09 that explained most of the variation for seed color and meal quality traits (Rahman and McVetty 2011; Liu et al. 2012; Kebede et al. 2012).

Significant genetic variation for phytate content has been reported across 505 genotypes of rapeseed. It ranged from 0.41 to 0.97 mg/5 seeds (Liu et al. 2021). A multidrug resistance-associated protein 5 (*BnaA9.MRP5*) gene was also identified as a candidate gene, with eight distinct haplotypes in association with both seed phytate content and concentration using GWAS in 505 accessions. RNAi-mediated downregulation of *MIPS* genes has been known to decrease phytate amount in a patent (Georges et al. 2006). Phytic acid mutants with knockout mutations in the six genes *BnMIPS, BnMIK, Bn2-PGK*, *BnIPK1, BnIPK2*, and *BnMRP5* were identified in an EMS (ethyl methane sulfonate) mutagenized population of rapeseed after implementation of high-throughput NGS screening protocol (Sashidhar et al. 2020). The study also identified double mutants of *Bn.2-PGK2*, which revealed substantial decrease in phytic acid content. In order to decrease the phytic acid content in rapeseed, three functional paralogs of an important enzyme

BnITPK (inositol tetrakisphosphate kinase) were knocked out using CRISPR-Cas9 mutagenesis in spring rapeseed cultivar Haydn. The mutants with low phytic acid and presence of free phosphorous were obtained.

7.3 Seed Storage Protein (SSP) Content and Composition

Substantial genetic variation for protein content and quality has been observed in rapeseed germplasm (Schatzki et al. 2014; Stolte et al. 2022). The content of cruciferin and napin ranged from 25.5 to 34.7 μg/sample and from 20.0 to 26.4 μg/sample, respectively, across 30 winter rapeseed hybrids of canola quality (Stolte et al. 2022). However, a few genomic studies have been carried out to reveal the molecular regulation of seed storage protein content and its composition. The outcomes of those studies revealed common or closely linked QTLs for seed oil and protein content but controlling the traits in an opposite manner. This is usually expected as both oil and protein compete for common substrates in the metabolic pathway, thus must be partly regulated by the same genes. Also, seed oil is significantly negatively correlated to protein quantity (Gül et al. 2003). *BnFus3* mutant seeds of *B. napus* depicted declined level of oil proportion with increased protein content. Decreased expression of sucrose photo-assimilation and glycolysis pathway related genes was observed in the seeds. This might have affected the synthesis of both oil and protein to alter their content in the seeds (Elahi et al. 2015). To identify the unique/separate QTLs for oil and protein content, Zhao et al. (2006) used a conditional mapping approach. The study detected five QTLs localized on chromosome A07, A09, C01, C08, and C09 for protein synthesis uncommon to QTLs for oil content (Zhao et al. 2006). Three QTLs for napin were envisioned on chromosomes A02, C06, and C09, and two QTLs for cruciferin content were observed on A02 and C19 (Schatzki et al. 2014). Four QTLs for protein content of defatted meal were found on chromosomes A01, A07, and C03. Of these, QTL located on chromosome A01 was associated to five traits, i.e., PA, OA, LA, oil content, and protein content. The variation in the region resulted in decreased PA, LA, and protein content in the defatted meal and increased oil content in the seed (Teh and Mollers 2016). Potential candidate genes encoding seed storage 2S, caleosin, oleosin, and cruciferin were identified for 38 QTLs regulating seed protein content in *B. napus* by Chao et al. (2017). Some promising attempts have also been undertaken to engineer rapeseed with more desirable levels of napin through introducing a Brazil nut (*Bertholletia excelsa*) 2S gene (Guerche et al. 1990, Altenbach et al. 1992) or expression of an antisense gene for cruciferin (Kohno-Murase et al. 1994; Hannoufa et al. 2014). In these cases, transgenic plants showed increased levels of lysine, cysteine, and methionine essential amino acids which led to more napins in their seeds. Also, the increase in napin content was counterbalanced by a decrease in cruciferins, thus suggesting that the 12S/2S ratio is tightly regulated. The lysine content of canola meal protein has also been improved by disrupting the feedback regulation of lysine during biosynthesis (Falco et al. 1995).

References

Abidi SL, List GR, Rennick KA (1999) Effect of genetic modification on the distribution of minor constituents in canola oil. J Am Oil Chem Soc 76:463–467

Alashi AM, Blanchard CL, Mailer RJ, Agboola SO, Mawson AJ, He R, Girgih A, Aluko RE (2014) Antioxidant properties of Australian canola meal protein hydrolysates. Food Chem 146:500–506

Altenbach SB, Kuo CC, Staraci LC, Pearson KW, Wainwright C, Georgescu A, Townsend J (1992) Accumulation of a *Brazil* nut albumin in seeds of transgenic canola results in enhanced levels of seed protein methionine. Plant Mol Biol 18:235–245

Amar S (2007) Genetic variation and inheritance of phytosterol content in *Brassica napus* L. Ph.D. Thesis, University of Gottingen, Germany

Amar S, Becker HC, Mollers C (2009) Genetic variation in phytosterol content of winter rapeseed (*Brassica napus* L.) and development of NIRS calibration equations. Plant Breed 128:78–83

Amar S, Ecke W, Becker HC, Mollers C (2008) QTL for phytosterol and sinapate ester content in *Brassica napus* L. collocate with the two erucic acid genes. Theor Appl Genet 116:1051–1061

Appelhagen I, Thiedig K, Nordholt N, Schmidt N, Huep G, Sagasser M, Weisshaar B (2014) Update on transparent testa mutants from *Arabidopsis thaliana*: characterization of new alleles from an isogenic collection. Planta 240:955–970

Augustine R, Mukhopadhyay A, Bisht NC (2013) Targeted silencing of BjMYB28 transcription factor gene directs development of low glucosinolate lines in oilseed *Brassica juncea*. Plant Biotechnol J 11:855–866

Baaij BM, Kim HK, Grosser K, Worrich A, de Jong TJ (2018) Slug herbivory on hybrids of the crop *Brassica napus* and its wild relative *B. rapa*. Basic Appl Ecol 31:52–60

Bai S, Engelen S, Denolf P, Wallis JG, Lynch K, Bengtsson JD, Thournout MV, Haesendonckx B, Browse J (2019) Identification, characterization and field testing of *Brassica napus* mutants producing high–oleic oils. Plant J 98:33–41

Bao B, Chao H, Wang H, Zhao W, Zhang L, Raboanatahiry N, Wang X, Wang B, Jia H, Li M (2018) Stable, environmental specific and novel QTL identification as well as genetic dissection of fatty acid metabolism in *Brassica napus*. Front Plant Sci 9:1018

Barret P, Delourme R, Brunel D, Jourdren C, Horvais R, Renard M (1999) Low linolenic acid level in rapeseed can be easily assessed through the detection of two single base substitution in Fad3 genes. In: Proceedings of the 10th international rapeseed congress. Canberra, Australia, pp. 26–29

Baskar V, Park SW (2015) An update on potential perspectives of glucosinolates on protection against microbial pathogens and endocrine dysfunctions in humans. Crit Rev Food Sci Nutr 56: 2231–2249

Behnke N, Suprianto E, Möllers CA (2018) Major QTL on chromosome C05 significantly reduces acid detergent lignin (ADL) content and increases seed oil and protein content in oilseed rape (*Brassica napus* L.). Theor Appl Genet 131:2477–2492

Blazevic I, Montaut S, Burcul F, Olsen CE, Burow M, Rollin P, Agerbirk N (2020) Glucosinolate structural diversity, identification, chemical synthesis and metabolism in plants. Phytochemistry 169:112100

Branca F, Ferrari M (2002) Impact of micronutrient deficiencies on growth: the stunting syndrome. Ann Nutr Metab 46:8–17

Cabral CE, Klein MRST (2017) Phytosterols in the treatment of hypercholesterolemia and prevention of cardiovascular diseases. Arq Bras Cardiol 109:475–482

Cao Z, Tian F, Wang N, Jiang C, Lin B, Xia W, Shi J, Long Y, Zhang C, Meng J (2010) Analysis of QTLs for erucic acid and oil content in seeds on A8 chromosome and the linkage drag between the alleles for the two traits in *Brassica napus*. J Genet Genom 37:231–240

Celenza JL, Quiel JA, Smolen GA, Merrikh H, Silvestro AR, Normanly J, Bender J (2005) The Arabidopsis *ATR1*Myb transcription factor controls indolic glucosinolate homeostasis. Plant Physiol 137:253–262

Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Correa M (2014) Early allopolyploid evolution in the post–Neolithic *Brassica napus* oilseed genome. Science 345:950–953

Chao H, Guo L, Zhao W, Li H, Li M (2022) A major yellow–seed QTL on chromosome A09 significantly increases the oil content and reduces the fiber content of seed in *Brassica napus*. Theor Appl Genet 135:1293–1305

Chao H, Wang H, Wang X, Guo L, Gu J, Zhao W, Li B, Chen D, Raboanatahiry N, Li M (2017) Genetic dissection of seed oil and protein content and identification of networks associated with oil content in *Brassica napus*. Sci Rep 7:1–16

Charlton KM, Corner AH, Davey K, Kramer JK, Mahadevan S, Sauer FD (1975) Cardiac lesions in rats fed rapeseed oils. Can J Comp Med 39:261–269

Chauhan JS, Singh KH, Singh VV, Kumar S (2010) Hundred years of rapeseed–mustard breeding in India: accomplishments and future strategies. Indian J Agric Sci 81:1093–1109

Chavadej S, Brisson N, McNeil JN, De Luca V (1994) Redirection of tryptophan leads to production of low indole glucosinolate canola. Proc Natl Acad Sci U S A 91:2166–2170

Ding LN, Guo XJ, Li M, Fu ZL, Yan SZ, Zhu KM, Wang Z, Tan XL (2019) Improving seed germination and oil contents by regulating the GDSL transcriptional level in *Brassica napus*. Plant Cell Rep 38:243–253

Elahi N, Duncan RW, Stasolla C (2015) Decreased seed oil production in FUSCA3 *Brassica napus* mutant plants. Plant Physiol Biochem 96:222–230

Elahi N, Duncan RW, Stasolla C (2016) Modification of oil and glucosinolate content in canola seeds with altered expression of *Brassica napus* LEAFY COTYLEDON1. Plant Physiol Biochem 100:52–63

Elhiti M, Yang C, Chan A, Durnin DC, Belmonte MF, Ayele BT, Tahir M, Stasolla C (2012) Altered seed oil and glucosinolate levels in transgenic plants overexpressing the *Brassica napus* SHOOTMERISTEMLESS gene. J Exp Bot 63:4447–4461

Endrigkeit J, Wang X, Cai D, Zhang C, Long Y, Meng J, Jung C (2009) Genetic mapping, cloning, and functional characterization of the BnaX. VTE4 gene encoding a γ–tocopherol methyltransferase from oilseed rape. Theor Appl Genet 119:567–575

Falco SC, Guida T, Locke M, Mauvais J, Sanders C, Ward RT, Webber P (1995) Transgenic canola and soybean seeds with increased lysine. Nat Biotechnol 13:577–582

Falentin C, BreEgeon M, Lucas MO, Deschamps M, Leprince F, Fournier MT, Delourme R, Renard M (2007) Identification of Fad2 mutations and development of allele–specific markers for high oleic acid content in rapeseed (*Brassica napus* L.). In: Proceedings of the 12th international rapeseed congress. Wuhan, China, pp 117–119

Feng J, Long Y, Shi L, Shi JQ, Barker G, Meng JL (2012) Characterization of metabolite quantitative trait loci and metabolic networks that control glucosinolate concentration in the seeds and leaves of *Brassica napus*. New Phytol 193:96–108

Fieldsen J, Milford GFJ (1994) Changes in glucosinolates during crop development in single- and double-low genotypes of winter oilseed rape (*Brassica napus*): I. production and distribution in vegetative tissues and developing pods during development and potential role in the recycling. Ann Appl Biol 124:531–542

Finlayson AJ, Krzymanski J, Downey RK (1973) Comparison of chemical and agronomic characteristics of two *Brassica napus* L. cultivars, Bronowski and target. J Am Oil Chem Soc 50:407–410

Fourmann M, Barret P, Renard M, Pelletier G, Delourme R, Brunel D (1998) The two genes homologous to Arabidopsis *FAE1* co–segregate with the two loci governing erucic acid content in *Brassica napus*. Theor Appl Genet 96:852–858

Frerigmann H, Gigolashvili T (2014) MYB34, MYB51 and MYB122 distinctly regulate Indolic Glucosinolate biosynthesis in *Arabidopsis thaliana*. Mol Plant 7:814–828

Friedt W, Snowdon RJ (2010) Oilseed rape. In: Vollmann J, Rajcan I (eds) Handbook of plant breeding– oil crops. Springer, New York, pp 91–126

Fritsche S, Wang X, Li J, Stich B, Kopisch-Obuch FJ, Endrigkeit J, Leckband G, Dreyer FW, Meng J, Jung C (2012) A candidate gene–based association study of tocopherol content and composition in rapeseed (*Brassica napus*). Front Plant Sci 3:129

Fu Y, Mason AS, Zhang Y, Yu H (2021) Identification and development of KASP markers for novel mutant BnFAD2 alleles associated with elevated oleic acid in *Brassica napus*. Front Plant Sci 12: 715633

Gacek K, Bayer PE, Anderson R, Severn-Ellis AA, Wolko J, Łopatynska A, Matuszczak M, Bocianowski J, Edwards D, Batley J (2021) QTL genetic mapping study for traits affecting meal quality in winter oilseed rape (*Brassica napus* L.). Genes 12:1235

Gacek K, Bayer PE, Bartkowiak-Broda I, Szala L, Bocianowski J, Edwards D, Batley J (2017) Genome–wide association study of genetic control of seed fatty acid biosynthesis in *Brassica napus*. Front Plant Sci 7:2062

Georges F, Hussain AA, Keller WA (2006) Method for reducing phytate in canola meal using genetic manipulation involving myo–inositol 1–phospathe synthase gene. US patent 7,148,064 B1, 12 Dec 2006

Gigolashvili T, Berger B, Flügge UI (2008a) Specific and coordinated control of indolic and aliphatic glucosinolate biosynthesis by R2R3–MYB transcription factors in *Arabidopsis thaliana*. Phytochem Rev 8:3–13

Gigolashvili T, Engqvist M, Yatusevich R, Muller C, Flugge UI (2008b) *HAG2/MYB76* and *HAG3/MYB29* exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in *Arabidopsis thaliana*. New Phytol 177:627–642

Gigolashvili T, Yatusevich R, Berger B, Muller C, Flugge UI (2007) The R2R3–MYB transcription factor HAG1/MYB28 is a regulator of methionine–derived glucosinolate biosynthesis in *Arabidopsis thaliana*. Plant J 51:247–261

Goffman FD, Becker HC (2002) Genetic variation of tocopherol content in a germplasm collection of *Brassica napus* L. Euphytica 125:189–196

Gordon MH, Miller LA (1997) Development of steryl ester analysis for the detection of admixtures of vegetable oils. J Am Oil Chem Soc 74:505–510

Gründemann C, Huber R (2018) Chemoprevention with isothiocyanates–from bench to bedside. Cancer Lett 414:26–33

Guan M, Huang X, Xiao Z, Jia L, Wang S, Zhu M, Qiao C, Wei L, Xu X, Liang Y, Wang R (2019) Association mapping analysis of fatty acid content in different ecotypic rapeseed using mrMLM. Front Plant Sci 9:1872

Guerche P, Dealmeida ERP, Schwarztein MA, Gander E, Krebbers E, Pelletier G (1990) Expression of the 2S albumin from bertholletia–excelsa in *Brassica napus*. Mol Genet Genomics 221:306–314

Gugliandolo A, Bramanti P, Mazzon E (2017) Role of vitamin E in the treatment of Alzheimer's disease: evidence from animal models. Int J Mol Sci 18:2504

Gul M, Becker HC, Ecke W (2003) QTL mapping and analysis of QTL x nitrogen interactions for protein and oil contents in *Brassica napus* L. In: Proceedings of the 11th international rapeseed congress. Copenhagen, Denmark, pp. 91–99

Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. Annu Rev Plant Biol 57:303–333

Han J, Luhs W, Sonntag K, Zahringer U, Borchardt DS, Wolter FP, Heinz E, Frentzen M (2001) Functional characterization of β–ketoacyl–CoA synthase genes from *Brassica napus* L. Plant Mol Biol 46:229–239

Hannoufa A, Pillai BVS, Chellamma S (2014) Genetic enhancement of *Brassica napus* seed quality. Transgenic Res 23:39–52

Harper AL, Trick M, Higgins J, Fraser F, Clissold L, Wells R, Hattori C, Werner P, Bancroft I (2012) Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. Nat Biotechnol 30:798

He Y, Yang Z, Tang M, Yang Y, Zhang Y, Liu S (2022) Enhancing canola breeding by editing a glucosinolate transporter gene lacking natural variation. Plant Physiol 188:1848–1851

Hirai MY, Sugiyama K, Sawada Y, Tohge T, Obayashi T, Suzuki A, Araki R, Sakurai N, Suzuki H, Aoki K (2007) Omics–based identification of Arabidopsis *Myb* transcription factors regulating aliphatic glucosinolate biosynthesis. Proc Natl Acad Sci U S A 104:6478–6483

Howell PM, Sharpe AG, Lydiate DJ (2003) Homoeologous loci control the accumulation of seed glucosinolates in oilseed rape (*Brassica napus*). Genome 46:454–460

Hu X, Sullivan-Gilbert M, Gupta M, Thompson SA (2006) Mapping of the loci controlling oleic and linolenic acid contents and development of *fad2* and *fad3* allele–specific markers in canola (*Brassica napus* L.). Theor Appl Genet 113:497–507

Huang H, Cui T, Zhang L, Yang Q, Yang Y, Xie K, Fan C, Zhou Y (2020) Modifications of fatty acid profile through targeted mutation at *BnaFAD2* gene with CRISPR/Cas9–mediated gene editing in *Brassica napus*. Theor Appl Genet 133:2401–2411

Jhingan S, Harloff HJ, Abbadi A, Welsch C, Blümel M, Tasdemir D, Jung C (2023) Reduced glucosinolate content in oilseed rape (*Brassica napus* L.) by random mutagenesis of *BnMYB28* and *BnCYP79F1* genes. Sci Rep 13(1):2344

Jiang C, Shi J, Li R, Long Y, Wang H, Li D, Zhao J, Meng J (2014) Quantitative trait loci that control the oil content variation of rapeseed (*Brassica napus* L.). Theor Appl Genet 127:957–968

Jourdren C, Barret P, Brunel D, Delourme R, Renard M (1996) Specific molecular marker of the genes controlling linolenic acid content in rapeseed. Theor Appl Genet 93:512–518

Kaiser F, Harloff HJ, Tressel RP, Kock T, Schulz C (2021) Effects of highly purified rapeseed protein isolate as fishmeal alternative on nutrient digestibility and growth performance in diets fed to rainbow trout (*Oncorhynchus mykiss*). Aquac Nutr 27:1352–1362

Karunarathna NL, Wang H, Harloff HJ, Jiang L, Jung C (2020) Elevating SEED oil content in a polyploid crop by induced mutations in *SEED FATTY ACID REDUCER* genes. Plant Biotechnol J 18:2251–2266

Katavic V, Mietkiewska E, Barton DL, Giblin EM, Reed DW, Taylor DC (2002) Restoring enzyme activity in nonfunctional low erucic acid *Brassica napus fatty acid elongase 1* by a single amino acid substitution. Eur J Biochem 269:5625–5631

Kebede B, Cheema K, Greenshields DL, Li C, Selvaraj G, Rahman H (2012) Construction of genetic linkage map and mapping of QTL for seed color in *Brassica rapa*. Genome 55:813–823

Kim YB, Li XH, Kim SJ, Kim HH, Lee J, Kim HR, Park SU (2013) *MYB* transcription factors regulate glucosinolate biosynthesis in different organs of Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). Molecules 18:8682–8695

Kittipol V, He Z, Wang L, Adams TD, Langer S, Bancroft I (2019) Genetic architecture of glucosinolate variation in *Brassica napus*. J Plant Physiol 240:152988

Kohno-Murase J, Murase M, Ichikawa H, Imamura J (1994) Effects of an antisense napin gene on seed storage compounds in transgenic *Brassica napus* seeds. Plant Mol Biol 26:1115–1124

Kondra ZP, Stefansson BR (1970) Inheritance of major glucosinolates of rapeseed (*Brassica napus*) meal. Can J Plant Sci 50:643–647

Kumar R, Raclaru M, Schüßeler T, Gruber J, Sadre R, Lühs W, Zarhloul KM, Friedt W, Enders D, Frentzen M, Weier D (2005) Characterisation of plant tocopherol cyclases and their over-expression in transgenic *Brassica napus* seeds. FEBS Lett 579:1357–1364

Lassner MW, Lardizabal K, Metz JG (1996) A jojoba beta–Ketoacyl–CoA synthase cDNA complements the canola fatty acid elongation mutation in transgenic plants. Plant Cell 8:281–292

Lee KR, Kim EH, Roh KH, Kim JB, Kang HC, Go YS, Suh MC, Kim HU (2016) High-oleic oilseed rapes developed with seed–specific suppression of *FAD2* gene expression. Appl Biol Chem 59:669–676

Li F, Chen B, Xu K, Wu J, Song W, Bancroft I, Harper AL, Trick M, Liu S, Gao G, Wang N, Yan G, Qiao J, Li J, Li H, Xiao X, Zhnag T, Wu X (2014) Genome–wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (*Brassica napus* L.). DNA Res 21:355–367

Lian J, Lu X, Yin N, Ma L, Jing L, Xue L, Li J, Lu J, Lei B, Wang R, Chai Y (2016) Silencing of *BnTT1* family genes affects seed flavonoid biosynthesis and alters seed fatty acid composition in *Brassica napus*. Plant Sci 254:32–47

Lickfett T, Matthäus B, Velasco L, Möllers C (1999) Seed yield, oil and phytate concentration in the seeds of two oilseed rape cultivars as affected by different phosphorus supply. Eur J Agron 11: 293–299

Lipşa F, Snowdon R, Friedt W (2012) Quantitative genetic analysis of condensed tannins in oilseed rape meal. Euphytica 184:195–205

Liu H, Li X, Zhang Q, Yuan P, Liu L, King GJ, Ding G, Wang S, Cai H, Wang C, Xu F, Shi L (2021) Integrating a genome-wide association study with transcriptomic data to predict candidate genes and favourable haplotypes influencing *Brassica napus* seed phytate. DNA Res 28:1–11

Liu H, Lin B, Ren Y, Hao P, Huang L, Xue B, Jiang L, Zhu Y, Hua S (2022b) CRISPR/Cas9–mediated editing of double loci of *BnFAD2* increased the seed oleic acid content of rapeseed (*Brassica napus* L.). Front Plant Sci 13:1034215

Liu J, Hao W, Liu J, Fan S, Zhao W, Deng L, Wang X, Hu Z, Hua W, Wang H (2019) A novel chimeric mitochondrial gene confers cytoplasmic effects on seed oil content in polyploid rapeseed (*Brassica napus*). Mol Plant 12:582–596

Liu L, Stein A, Wittkop B, Sarvari P, Li J, Yan X, Dreyer F, Frauen M, Friedt W, Snowdon RJ (2012) A knockout mutation in the lignin biosynthesis gene CCR1 explains a major QTL for acid detergent lignin content in *Brassica napus* seeds. Theor Appl Genet 124:1573–1586

Liu Y, Du Z, Lin S, Li H, Lu S, Guo L, Tang S (2022a) CRISPR/Cas9–targeted mutagenesis of *BnaFAE1* genes confers low–Erucic acid in *Brassica napus*. Front Plant Sci 13:848723

Liu Y, Zhou X, Yan M, Wang P, Wang H, Xin Q, Yang L, Hong D, Yang G (2020a) Fine mapping and candidate gene analysis of a seed glucosinolate content QTL, qGSL–C2, in rapeseed (*Brassica napus* L.). Theor Appl Genet 133:479–490

Liu Z, Hammerlindl J, Keller W, McVetty P, Daayf F, Quiros C, Li G (2011) *MAM* gene silencing leads to the induction of C3 and reduction of C4 and C5 side–chain aliphatic glucosinolates in *Brassica napus*. Mol Breed 27:467–478

Liu S, Huang H, Yi X, Zhang Y, Yang Q, Zhang C, Fan C, Zhou Y (2020b) Dissection of genetic architecture for glucosinolate accumulations in leaves and seeds of *Brassica napus* by genome–wide association study. Plant Biotechnol J 18:1472–1484

Long Y, Wang J, Wang YY, Zhang JW, Wang J, Pei XW (2016) Comparative analysis of *MYB28* homologs and development of a MYB28–specific marker in *Brassica napus* L. Mol Breed 36:126

Lu G, Harper AL, Trick M, Morgan C, Fraser F, O'Neill C, Bancroft I (2014) Associative transcriptomics study dissects the genetic architecture of seed glucosinolate content in *Brassica napus*. DNA Res 21:613–625

Magrath R, Bano F, Morgner M, Parkin I, Sharpe A, Lister C, Dean C, Turner J, Lydiate D, Mithen R (1994) Genetics of aliphatic glucosinolates. I. Side chain elongation in *Brassica napu*s and *Arabidopsis thaliana*. Heredity 72:290

Malabat C, Atterby H, Chaudhry Q, Renard M, Gueguen J (2003) Genetic variability of rapeseed protein composition. In: Proceedings of the 11th international rapeseed congress. Copenhagen, Denmark, pp 1–3

Miao LY, Chao HB, Chen L, Wang H, Zhao W, Li B, Zhang L, Li H, Wang B, Li M (2019) Stable and novel QTL identification and new insights into the genetic networks affecting seed fiber traits in *Brassica napus*. Theor Appl Genet 132:1761–1775

Mitreiter S, Gigolashvili T (2021) Regulation of glucosinolate biosynthesis. J Exp Bot 72:70–91

Naczk M, Amarowicz R, Sullivan A, Shahidi F (1998) Current research developments on polyphenolics of rapeseed/canola: a review. Food Chem 62:489–502

Nesi N, Delourme R, Bregeon M, Falentin C, Renard M (2008) Genetic and molecular approaches to improve nutritional value of *Brassica napus* L. seed. C R Biol 331(10):763–771

Nour-Eldin HH, Andersen TG, Burow M, Madsen SR, Jørgensen ME, Olsen CE, Dreyer I, Hedrich R, Geiger D, Halkier BA (2012) NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds. Nature 488:531–534

Nour-Eldin HH, Halkier BA (2009) Piecing together the transport pathway of aliphatic glucosinolates. Phytochem Rev 8:53–67

Nowak H, Kujawa K, Zadernowski R, Roczniak B, KozŁowska H (1992) Antioxidative and bactericidal properties of phenolic compounds in rapeseeds. Eur J Lipid Sci Technol 94:149–152

Okuzaki A, Ogawa T, Koizuka C, Kaneko K, Inaba M, Imamura J, Koizuka N (2018) CRISPR/Cas9–mediated genome editing of the fatty acid desaturase 2 gene in *Brassica napus*. Plant Physiol Biochem 131:63–69

Parkin I, Magrath R, Keith D, Sharpe A, Mithen R, Lydiate D (1994) Genetics of aliphatic glucosinolates. II. Hydroxylation of alkenyl glucosinolates in *Brassica napus*. Heredity 72:594

Peng Q, Hu Y, Wei R, Zhang Y, Guan C, Ruan Y, Liu C (2010) Simultaneous silencing of *FAD2* and *FAE1* genes affects both oleic acid and erucic acid contents in *Brassica napus* seeds. Plant Cell Rep 29:317–325

Petrie JR, Zhou XR, Leonforte A, McAllister J, Shrestha P, Kennedy Y, Belide S, Buzza G, Gororo N, Gao W, Lester G (2020) Development of a *Brassica napus* (canola) crop containing fish oil–like levels of DHA in the seed oil. Front Plant Sci 11:727

Piironen V, Lindsay DG, Miettinen TA, Toivo J, Lampi AM (2000) Plant sterols: biosynthesis, biological function and their importance to human nutrition. J Sci Food Agric 80:939–966

Qu C, Jia L, Fu F, Zhao H, Lu K, Wei L, Xu X, Liang Y, Li S, Wang R, Li J (2017) Genome–wide association mapping and identification of candidate genes for fatty acid composition in *Brassica napus* L. using SNP markers. BMC Genomics 18:1–17

Raclaru M, Gruber J, Kumar R, Sadre R, Luehs W, Zarhloul MK, Friedt W, Frentzen M, Weier D (2006) Increase of the tocochromanol content in transgenic *Brassica napus* seeds by over-expression of key enzymes involved in prenylquinone biosynthesis. Mol Breed 18:93–107

Rahman M, Li G, Schroeder D, McVetty PBE (2010) Inheritance of seed coat color genes in *Brassica napus* (L.) and tagging the genes using SRAP, SCAR and SNP molecular markers. Mol Breed 26:439–453

Rahman M, McVetty PBE (2011) A review of *brassica* seed color. Can J Plant Sci 91:437–446

Russo M, Yan F, Stier A, Klasen L, Honermeier B (2021) Erucic acid concentration of rapeseed (*Brassica napus* L.) oils on the German food retail market. Food Sci Nutr 9:3664–3672

Sashidhar N, Harloff HJ, Jung C (2020) Identification of phytic acid mutants in oilseed rape (Brassica napus) by large–scale screening of mutant populations through amplicon sequencing. New Phytol 225:2022–2034

Scarth R, Rimmer SR, McVetty PBE (1995) Apollo low linolenic summer rape. Can J Plant Sci 75: 203–204

Scarth RPBE, McVetty PBE, Rimmer SR, Stefansson BR (1988) Stellar low linolenic–high linoleic acid summer rape. Can J Plant Sci 68:509–511

Schatzki J, Ecke W, Becker HC, Mollers C (2014) Mapping of QTL for the seed storage proteins cruciferin and napin in a winter oilseed rape doubled haploid population and their inheritance in relation to other seed traits. Theor Appl Genet 127:1213–1222

Schneider C (2005) Chemistry and biology of vitamin E. Mol Nutr Food Res 49:7–30

Shahidi F, Naczk M (1992) An overview of the phenolics of canola and rapeseed: chemical, sensory and nutritional significance. J Am Oil Chem Soc 69:917–924

Shi J, Lang C, Wang F, Wu X, Liu R, Zheng T, Zhang D, Chen J, Guanting W (2017) Depressed expression of *FAE1* and *FAD2* genes modifies fatty acid profiles and storage compounds accumulation in *Brassica napus* seeds. Plant Sci 263:177–182

Sonderby IE, Geu-Flores F, Halkier BA (2010) Biosynthesis of glucosinolates–gene discovery and beyond. Trends Plant Sci 15:283–290

Song JM, Guan Z, Hu J, Guo C, Yang Z, Wang S, Liu D, Wang B, Lu S, Zhou R, Xie WZ (2020) Eight high–quality genomes reveal pan–genome architecture and ecotype differentiation of *Brassica napus*. Nat Plants 6:34–45

Spasibionek S, Mikołajczyk K, Piętka T, Krótka K, Matuszczak M, Nowakowska J, Michalski K, Bartkowiak-Broda I (2020) Marker assisted selection of new high oleic and low linolenic winter oilseed rape (*Brassica napus* L.) inbred lines revealing good agricultural value. PLoS One 15: 0233959

Stein A, Wittkop B, Liu L, Obermeier C, Friedt W, Snowdon RJ (2013) Dissection of a major QTL for seed colour and fibre content in *Brassica napus* reveals colocalization with candidate genes for phenylpropanoid biosynthesis and flavonoid deposition. Plant Breed 132:382–389

Stolte N, Vettel J, Möllers C (2022) Genetic variation for seed storage protein composition in rapeseed (*Brassica napus*) and development of near–infrared reflectance spectroscopy calibration equations. Plant Breed 141:408–417

Swati, Sehwag S, Das M (2015) A brief overview: present status on utilization of mustard oil and cake. Indian J Tradit Knowl 14:244–250

Tan H, Yang X, Zhang F, Zheng X, Qu C, Mu J, Fu F, Li J, Guan R, Zhang H, Wang G (2011) Enhanced seed oil production in canola by conditional expression of *Brassica napus* leafy cotyledon1 and LEC1–like in developing seeds. Plant Physiol 156:1577–1588

Tang S, Zhao H, Lu S, Yu L, Zhang G, Zhang Y, Yang QY, Zhou Y, Wang X, Ma W, Xie W (2021) Genome–and transcriptome–wide association studies provide insights into the genetic basis of natural variation of seed oil content in *Brassica napus*. Mol Plant 14:470–487

Tangtaweewipat S, Cheva-Isarakul B, Sangsrijun P (2004) The use of mustard meal as a protein source in broiler diets. Songklanakarin J Sci Technol 26:23–30

Taylor DC, Katavic V, Zou J, MacKenzie SL, Keller WA, An J, Friesen W, Barton DL, Pedersen KK, Michael Giblin E, Ge Y (2002) Field testing of transgenic rapeseed cv. Hero transformed with a yeast sn–2 acyltransferase results in increased oil content, erucic acid content and seed yield. Mol Breed 8:317–322

Teh L, Mollers C (2016) Genetic variation and inheritance of phytosterol and oil content in a doubled haploid population derived from the winter oilseed rape Sansibar× Oase cross. Theor Appl Genet 129:181–199

Uzunova M, Ecke W, Weissleder K, RoBbelen G (1995) Mapping the genome of rapeseed (*Brassica napus* L.). I. Construction of an RFLP linkage map and localization of QTLs for seed glucosinolate content. Theor Appl Genet 90:194–204

Velasco P, Soengas P, Vilar M, Cartea ME, del Rio M (2008) Comparison of glucosinolate profiles in leaf and seed tissues of different *Brassica napus* crops. J Am Soc Hortic Sci 133:551–558

Vigeolas H, Waldeck P, Zank T, Geigenberger P (2007) Increasing seed oil content in oil-seed rape (*Brassica napus* L.) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seed-specific promoter. Plant Biotechnol J 5:431–441

Walsh TA, Bevan SA, Gachotte DJ, Larsen CM, Moskal WA, Merlo PA, Sidorenko LV, Hampton RE, Stoltz V, Pareddy D, Anthony GI (2016) Canola engineered with a microalgal polyketide synthase–like system produces oil enriched in docosahexaenoic acid. Nat Biotechnol 34: 881–887

Wanasundara JP, McIntosh TC, Perera SP, Withana-Gamage TS, Mitra P (2016) Canola/rapeseed protein–functionality and nutrition. OCl 23:D407

Wang B, Wu Z, Li Z, Zhang Q, Hu J, Xiao Y, Cai D, Wu J, King GJ, Li H, Liu K (2018) Dissection of the genetic architecture of three seed–quality traits and consequences for breeding in *Brassica napus*. Plant Biotechnol J 16:1336–1348

Wang H, Wang Q, Pak H, Yan T, Chen M, Chen X, Wu D, Jiang L (2021) Genome–wide association study reveals a patatin–like lipase relating to the reduction of seed oil content in *Brassica napus*. BMC Plant Biol 21:1–12

Wang N, Shi L, Tian F, Ning H, Wu X, Long Y, Meng J (2010) Assessment of FAE1 polymorphisms in three brassica species using EcoTILLING and their association with differences in seed erucic acid contents. BMC Plant Biol 10:1–11

Wang X, Zhang C, Li L, Fritsche S, Endrigkeit J, Zhang W, Long Y, Jung C, Meng J (2012) Unraveling the genetic basis of seed tocopherol content and composition in rapeseed (*Brassica napus* L.). Plos One 11:e50038

Wells R, Trick M, Soumpourou E, Clissold L, Morgan C, Werner P, Gibbard C, Clarke M, Jennaway R, Bancroft I (2014) The control of seed oil polyunsaturate content in the polyploid crop species *Brassica napus*. Mol Breed 33:349–362

Wen J, Xu J, Long Y, Xu H, Wu J, Meng J, Shi C (2015) Mapping QTLs controlling beneficial fatty acids based on the embryo and maternal plant genomes in *Brassica napus* L. J Am Oil Chem Soc 92:541–552

Wen J, Zhu L, Qi L, Ke H, Yi B, Shen J, Tu J, Ma C, Fu T (2012) Characterization of interploid hybrids from crosses between *Brassica juncea* and *B. oleracea* and the production of yellow–seeded *B. napus*. Theor Appl Genet 125:19–32

Wittkop B, Snowdon R, Friedt W (2009) Status and perspectives of breeding for enhanced yield and quality of oilseed crops for Europe. Euphytica 170:131–140

Wittstock U, Burow M (2010) Glucosinolate breakdown in Arabidopsis: mechanism, regulation and biological significance. The Arabidopsis Book 8:e0134

Wu G, Truksa M, Datla N, Vrinten P, Bauer J, Zank T, Cirpus P, Heinz E, Qiu X (2005) Stepwise engineering to produce high yields of very long–chain polyunsaturated fatty acids in plants. Nat Biotechnol 23:1013–1017

Wu G, Wu Y, Xiao L, Li X, Lu C (2008) Zero erucic acid trait of rapeseed (*Brassica napus* L.) results from a deletion of four base pairs in the fatty acid elongase 1 gene. Theor Appl Genet 116:491–499

Xiao Z, Zhang C, Tang F, Yang B, Zhang L, Liu J, Huo Q, Wang S, Li S, Wei L, Du H (2019) Identification of candidate genes controlling oil content by combination of genome–wide association and transcriptome analysis in the oilseed crop *Brassica napus*. Biotechnol Biofuels 12:1–16

Yan S, Li H, Chao H, He J, Ding Y, Zhao W, Zhang K, Xiong Y, Chen K, Zhang L, Li M (2022) Refinement of four major QTL for oil content in *Brassica napus* by integration of genome resequencing and transcriptomics. Crop J 10:627–637

Yang Q, Fan C, Guo Z, Qin J, Wu J, Li Q, Fu T, Zhou Y (2012) Identification of *FAD2* and *FAD3* genes in *Brassica napus* genome and development of allele–specific markers for high oleic and low linolenic acid contents. Theor Appl Genet 125:715–729

Yao M, Guan M, Zhenqian Z, Zhang Q, Cui Y, Chen H, Liu W, Jan H, Voss-Fels K, Werner C, He X, Liu Z, Guan C, Snowdon R, Hua W, Qian L (2019) GWAS and co-expression network combination uncovers multigenes with close linkage effects on oleic acid content accumulation in *Brassica napus*. BMC Genomics 21:1–12

Yin L, Chen HC, Cao BH, Lei JJ, Chen GJ (2017) Molecular characterization of MYB28 involved in aliphatic glucosinolate biosynthesis in Chinese kale (*Brassica oleracea* var. *alboglabra* bailey). Front Plant Sci 8:1083

Zafar S, Tang MQ, Wang YK, Sarwar R, Liu SY, Tan XL (2020) Candidate genes–association study to identify loci related to oleic acid in *Brassica napus* using SNP markers and their heterologous expression in yeast. Plant Physiol Biochem 146:294–302

Zhang D, Zhang Z, Unver T, Zhang B (2021) CRISPR/Cas: a powerful tool for gene function study and crop improvement. J Adv Res 29:207–221

Zhang K, Lu K, Qu C, Liang Y, Wang R, Chai Y, Li J (2013) Gene silencing of *BnTT10* family genes causes retarded pigmentation and lignin reduction in the seed coat of *Brassica napus*. PLoS One 8:e61247

Zhang K, Nie L, Cheng Q, Yin Y, Chen K, Qi F, Zou D, Liu H, Zhao W, Wang B, Li M (2019) Effective editing for lysophosphatidic acid acyltransferase 2/5 in allotetraploid rapeseed (*Brassica napus* L.) using CRISPR–Cas9 system. Biotechnol Biofuels 12:1–18

Zhao C, Xie M, Liang L, Yang L, Han H, Qin X, Zhao J, Hou Y, Dai W, Du C, Xiang Y (2022) Genome–wide association analysis combined with quantitative trait loci mapping and dynamic transcriptome unveil the genetic control of seed oil content in *Brassica napus* L. Front Plant Sci 13:929197

Zhao J, Huang J, Chen F, Xu F, Ni X, Xu H, Wang Y, Jiang C, Wang H, Xu A, Huang R (2012) Molecular mapping of *Arabidopsis thaliana* lipid–related orthologous genes in *Brassica napus*. Theor Appl Genet 124:407–421

Zhao JY, Becker HC, Zhang D, Zhang YF, Ecke W (2006) Conditional QTL mapping of oil content in rapeseed with respect to protein content and traits related to plant development and grain yield. Theor Appl Genet 113:33–38

Zhao Q, Wu J, Cai G, Yang Q, Shahid M, Fan C, Zhang C, Zhou Y (2019) A novel quantitative trait locus on chromosome A9 controlling oleic acid content in *Brassica napus*. Plant Biotechnol J 17:2313–2324

Zhu Q, King GJ, Liu X, Shan N, Borpatragohain P, Baten A, Wang P, Luo S, Zhou Q (2019) Identification of SNP loci and candidate genes related to four important fatty acid composition in *Brassica napus* using genome wide association study. PLoS One 14:e0221578

Zou J, Katavic V, Giblin EM, Barton DL, MacKenzie SL, Keller WA, Hu X, Taylor DC (1997) Modification of seed oil content and acyl composition in the brassicaceae by expression of a yeast sn–2 acyltransferase gene. Plant Cell 9:909–923

Nutragenomic Approaches in Sunflower: Genetic Improvement in Oil Quality

Manivannan Narayana and Ameena Premnath

Contents

M. Narayana (✉)
Centre of Excellence in Molecular Breeding, Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore, India
e-mail: nm68@tnau.ac.in

A. Premnath
Department of Plant Biotechnology, Centre for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore, India

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_14

Abstract

Sunflower is an important oilseed crop. Worldwide, it ranks fifth in cultivation among oilseed crops. Sunflower oil is preferred for domestic consumption as well as for cooking due to its nutritional quality. The sunflower oil quality is decided based on the proportion of various fatty acids and tocopherols. Recent breeding efforts in sunflower require a special focus on altering the oil quality. Breeding of new sunflower genotypes for food and nonfood industries assures a bright future for sunflower. Recent developments in molecular techniques helped to understand the genetic architecture of sunflower traits and the changes that happened during domestication. The availability of the sunflower genome sequence made a real breakthrough in sunflower molecular biology. This sequence could be effectively used to locate QTLs associated with various traits. It may help to understand the metabolic pathways of various quality traits. It may also help to replace SSR markers with SNPs. The valorization of the nutritional aspects of sunflower would throw the limelight in the realization of oil sustainability and enhance the circular economy in the global oilseed market.

Keywords

Sunflower · Genetic architecture · Oil quality · Fatty acid · Molecular biology · Genome sequence

1 Introduction

Sunflower (*Helianthus annuus* L.) is a diploid (2n = 34) annual plant. It belongs to the subtribe Helianthineae, tribe Helianthea, subfamily Asteroideae, and family Compositae (Asteraceae). The botanical name *Helianthus* comes from the Greek words *helios* (sun) and *anthos* (flower). The genus *Helianthus* is native to temperate North America. It has 14 annual and 37 perennial species. As per archaeological reports, the sunflower was cultivated by American Indians during 4625 BC (Jocic et al. 2015). Sunflower is used in many ways. It is used in the diet as nuts and flour. It is used as oil to decorate the hair and to protect the skin from the sun. It is also used to obtain colors (yellow and red). Sunflower is used for medical purposes and also as an ornamental plant.

Sunflower is an important oilseed crop. Globally, it ranked fifth among edible oilseed crop area with 22.9 million hectares. The nutritional value of vegetable oils is based on the proportion of various fatty acids (Fig. 1). The most important fatty acids of sunflower oil are linoleic acid (45–65%) (PUFA or poly unsaturated fatty acid) and oleic acid (20–30%) (MUFA or mono unsaturated fatty acid). In addition, palmitic and stearic acids are also available. Sunflower oils are categorized as high linoleic, high oleic, and mid-oleic. Mid-oleic sunflower oil has a minimum of 50–60% oleic acid. High oleic sunflower oil has 80–85% oleic acid. Oleic acid (18:1) only reduces cholesterol's atherogenic fraction. Therefore, oleic acid is

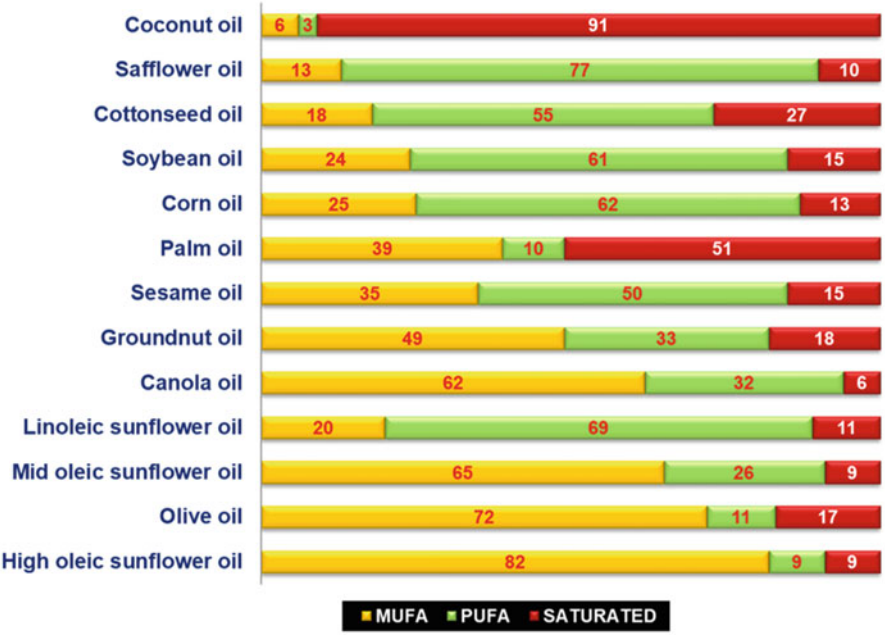


Fig. 1 Fatty acid profile of different vegetable oils

beneficial for cardiovascular disease prevention (Delplanque 2000). Dietary recommendations are the higher proportion of MUFA and reduction of saturated fats. Hence, the food manufacturing industry has more interest in high oleic sunflower oil. The qualitative manipulation of oils involves the changes in fatty acid composition. It decides the chemical properties of the oil and its use. Breeders exploit various breeding techniques to modify the proportion of the fatty acids.

Vegetable oils and fat are the third most important components in the human diet. Oil is the main cooking medium for majority of dishes. It increases the taste of those dishes. The suggested per capita oils and fat are 30 g per day. Hence, there is a need to enhance the oil quality to reduce the levels of bad (low density lipoprotein, "LDL") cholesterol.

2 Nutritional Composition of Sunflower

The common sunflower seed supplies many nutritious components like protein, unsaturated fats, fiber, vitamins (especially E), selenium, copper, zinc, folate, iron, and more. It can be used as cooking oil, enjoyed as a roasted or salted snack, dehulled, and included as a confectionary nut. Since sunflower seed is high in sulfuric amino acids, its meal is widely used as both livestock and pet feed. Sunflower oil quality is related to seed oil content and fatty acid composition and defines the oil's value for the industry. Sunflower seed contains 35–42% oil and is

naturally rich in linoleic acid (55–70%) and consequently poor in oleic acid (20–25%) (Premnath et al. 2016). Oleic acid is a mono unsaturated omega-9 fatty acid. It is capable of lowering triacyl glycerides and low density lipoprotein (LDL) cholesterol levels, increasing high density lipoprotein (HDL) cholesterol. Hence, it reduces the risk of heart attack (Guo et al. 2017).

Oleic and linoleic acids contribute almost 90% of the total fatty acid content in sunflower oil. High oleic sunflower oil is generally with more than 80% oleic acid. Linoleic sunflower oil has at least 65% linoleic acid. High oleic sunflower oil has a very neutral taste. It offers great stability without hydrogenation. Linoleic sunflower oil has susceptibility for oxidation, particularly during frying. The patents for high oleic seed (patent # 4627192) and high oleic oil (patent # 4743402) were issued on December 9, 1986, and May 10, 1988, respectively. These patents subsequently expired on December 9, 2003, and May 10, 2005, respectively, for high oleic seed and oil. Hence, the research work on high oleic oil has been started in the public research institutions as the expiry of the patent expired. The total tocopherol content in standard sunflower oil is 700–1000 mg/kg. Natural tocopherols are present in four isomers: α (5,7,8-trimethyltocol), β (5,8-dimethyltocol), γ (7,8-dimethyltocol), and δ (8-methyltocol). Standard sunflower oil contains mostly α-tocopherol (95%) and lower quantities of β-tocopherol (3%) and γ-tocopherol (2%) (Skoric et al. 2008). Tocopherols are considered to be natural antioxidants but have different *in vivo* and *in vitro* antioxidative activities. α-Tocopherol expresses maximum *in vivo* activity, also known as vitamin E activity, but low *in vitro* protection of extracted oil. γ-, δ-, and β-ocopherols are very important antioxidants, but with low vitamin E values (Packer and Jevic 2002).

Sunflower sprouts accumulate high amounts of caffeoylquinic acids (CQAs) including chlorogenic acid (5-CQA) and 1,5-diCQA. HaHQT2 was found to be the major isoform, which could be responsible for CQA biosynthesis during germination in both hypocotyls and cotyledons. Therefore, manipulation of this gene could be applied to biofortify sunflower sprouts as functional foods (Cheevarungnapakul et al. 2019).

Sunflowers have been employed in the preparation of various delicacies. Sunflower seeds can be processed into different forms, such as flour or roasted, baked, or boiled as composite functional foods. Sunflower remains a source of nutritional food for humans. Studies have revealed that sunflower seeds are rich in nutrients and certain different phytochemicals such as antioxidants, flavonols, phenolic acids, procyanidins, phytosterols, amino acids, dietary fiber, potassium, arginine monounsaturated, and polyunsaturated fatty acids which contribute to the improvement of human health (Adeleke and Babalola 2020). Ye et al. (2015) reported caffeic acid hexose I, caffeic acid hexose II, p-coumaric acid hexose, chlorogenic acid, isoquercitrin, 3,4-Di-*O*-caffeoylquinic acid, 1,5-Di-*O*-caffeoylquinic acid, 3,5-Di-*O*-caffeoylquinic acid, and 4,5-Di-*O*-caffeoylquinic acid as principal phenolic contents in sunflower florets. The presence of some essential amino acids such as aspartic acid, glutamic acid, serine, histidine, glycine, threonine, arginine, alanine, tyrosine, cysteine, valine, methionine, phenylalanine, isoleucine, leucine, lysine, and proline in sunflower products has also been reported (Karangwa et al. 2015). Sunflower seeds contain high amounts of vitamins like vitamin E, B, folate, and niacin and

minerals like calcium, copper, iron, magnesium, manganese, selenium, phosphorous, potassium, sodium, and zinc. By and large, the therapeutic potential of sunflower seeds has been proven to be medically curative for colds and coughs, as a substitute for quinine, exhibiting anti-malaria efficacy, and as a diuretic and expectorant (Islam et al. 2016). As a promising protein source, sunflower seeds in food preparation can be made as a substitute for soybean (Oliveira et al. 2019).

3 Genetic Resources of Health-Related (HR) Genes in Sunflower

A sustainable agricultural system and the security of global food supply are based on the maintenance of essential crop species and crop wild relatives (CWR). Gene banks' goals consist of encouraging the usage of CWR for crop improvement. It serves as the primary repository for the preservation of accessions and associated data and for distributing seeds for germplasm associated studies. Resources for germplasm can be categorized as *ex situ* or *in situ*. *In situ* resources are preserved in their natural settings in which they continue to change because of their surroundings. This is the best approach to keep populations alive, but in contrast to populations kept in gene banks, natural populations are frequently threatened by human activity. *Ex situ* conservation's most significant benefit is that it offers equal weight to resource availability and preservation.

3.1 Gene Pool Classification in Sunflower

The primary gene pool of sunflower contains cultivated and wild species of *H. annuus* and winter sunflower (*H. winteri* J. C. Stebbins). All these species are readily crossable (Stebbins et al. 2013). The secondary gene pool (e.g., *H. anomalus* S. F. Blake, *H. paradoxus* Heiser, *H. petiolaris*, and *H. deserticola* Heiser) consists of species that have undergone some degree of differentiation to the genome of cultivated species. These species have potential meiotic difficulties during hybridization. The tertiary germplasm pool (e.g., *H. hirsutus* Raf., *H. tuberosus*, and *H. divaricatus* L.) has a high degree of differentiation. Hence, it requires specialized techniques such as embryo rescue for the recovery of interspecific hybrids. Differentiation among the species can be measured through molecular, cytological, and morphological bases. The extent of wild species uses decreases from primary through tertiary pools due to differences in ploidy level and growth habit, as well as reproductive barriers (Warburton et al. 2017).

3.2 Crop Wild Relatives (CWRs) in Sunflower

Cultivated sunflower is a new crop when compared to the other major temperate crops. Although its genetic basis was restricted by a domestication bottleneck, the

huge number of sunflower CWR allows for the mining of a sizable genetic pool for crop improvement. Collection and preservation are made possible with little difficulty because North America is the principal region of origin for the genus *Helianthus*. There are 53 species in this genus. Of this, 14 and 39 are annuals and perennials, respectively (Stebbins et al. 2013). These species can be found from Canada to Mexico and alongside the Pacific and Atlantic coasts. Wild species constitute an important source of resistance genes due to the coevolution of sunflower CWR, their pathogens, and insect pests (Terzic et al. 2020). Information on wild species is highly useful to identify sources of tolerance genes in ecotypes (Seiler et al. 2017).

Sunflower CWR has high variation for tolerance to abiotic stresses. Salinity tolerance available in *Helianthus paradoxus* was incorporated into cultivated sunflower (Miller 1995). Hajjar and Hodgkin (2007) suggested that these saline tolerant genotypes could produce higher yield under salinity. CWR additionally offers the possibility of studying physiological processes of the survival mechanisms of species available under desert situation (Bowsher et al. 2016). *Helianthus argophyllus* has been significantly used for drought tolerance breeding. Baldini and Vannozzi (1998) stated that interspecific derivatives were obtained involving *H. argophyllus.* These derivatives were selected through divergent selection for physiological traits of *H. argophyllus.* These derivatives had better drought tolerant traits under drought situation than *H. annuus* lines. Resistance genes for rust, downy mildew, *Verticillium* wilt, *Alternaria* leaf spot, powdery mildew, *Phomopsis* stem canker and *Sclerotinia* wilt/rot, and broomrape were incorporated into cultivated sunflower from CWR (Seiler et al. 2017).

3.3 CWRs for Fatty Acid Composition in Sunflower

Two annual desert species, *Helianthus anomalus* Blake and *H. deserticola* Heiser, are excellent candidates for increasing oil concentration and enhancing quality based on their adaptation to desert environments. *H. anomalus* (sand sunflower) is a rare endemic species adapted to sand dune and swale habitats in Utah and northern Arizona (Heiser et al. 1969). Sand sunflower is a diploid annual species of hybrid origin that is endemic to active sand dunes, an extreme environment from its parents, *H. annuus* and *H. petiolaris* (Ludwig et al. 2004). *H. deserticola* (desert sunflower) is a xerophytic annual species found in sandy soils on the floor of the Great Basin Desert in small populations in western Nevada, in west central Utah, and along the border of Utah and Arizona (Heiser et al. 1969). *H. anomalus* has a larger seed and higher oil concentration than any other wild sunflower species. It also has the same chromosome number as cultivated sunflower. This will facilitate the introgression of agronomic traits from wild germplasm into cultivated sunflower. The lower saturated fatty acid profile in *H. anomalus* has the potential to reduce saturated fatty acids in cultivated sunflower. There appears to be sufficient variability in *H. anomalus* to introduce and select for high linoleic acid concentration and reduced saturated fatty acid concentrations in cultivated sunflower oil (Seiler 2007).

Reduced concentrations of saturated palmitic and stearic fatty acids have been observed in a population of wild *H. annuus* that had a combined concentration of 58 g kg^{-1}, and 65 g kg^{-1} was observed in a wild perennial species, *H. giganteus* (Seiler et al. 2017). These values are 50% lower than in the oil of cultivated sunflower (120 g kg^{-1}), providing new potential sources for reducing saturated fats in the oil.

The CWR of sunflower has played a vital role in the sunflower crop improvement. It will continue as one of the primary sources of genetic diversity for the sunflower crop. Traditional breeding technologies have established a basis for utilizing CWR in the improvement of cultivated sunflower. Incorporating CMS from wild sunflower facilitated the development of the globally valued sunflower crop. The incorporation of disease resistance and other biotic and abiotic traits has protected the investment in hybrid sunflower. Future developments are expected to take advantage of emerging technologies to increase the efficiencies of the breeding process of mining the genes from existing sunflower CWR.

4 Mode of Gene Action for High Oleic Acid Content in Sunflower

Genetic control of a high oleic mutation had been reported as one dominant gene *Ol* (Fick 1984; Urie 1985); major gene *Ol* and gene-modifier *ml* (Miller et al. 1987); recessive gene *ol* and dominant *Ml* (Martinez et al. 1999); three complementary genes *Ol1*, *Ol2*, and *Ol3* (Martinez et al. 1989); five genes *Ol1*, *Ol2*, *Ol3*, *Ol4*, and *Ol5* (Velasco et al. 2000); gene *Ol* with incomplete penetrance determined by genotypic epistatic factors of reversion (Demurin 2003); and high oleic locus *oleHOS* and suppressor locus *Su* (Lacombe et al. 2001; Berville 2010). Since the high oleic trait is such a complex one, these authors expressed the need for conducting further research with various crosses, locations, temperature, etc. to get a clear idea of the genetic system of oleic acid content (Dimitrijevic et al. 2017). The details on the genetics of high OAC in sunflower were presented in Table 1.

Genes responsible for tocopherol content of the seed were first reported in sunflower. Two independent genes, viz., *Tph1* and *Tph2*, were identified (Popov et al. 1988). *Tph1* controls the relationship among α- and β-tocopherol. *Tph2* gene controls the relationship among α- and γ-tocopherol. Both *Tph1* and *Tph2* genes under recessive homozygous condition have an epistatic effect. It leads to the production of δ-tocopherol (Demurin 1993). Velasco et al. (2004) reported that one recessive gene is responsible for β- and γ-tocopherol content. Demurin et al. (2004) and Vera-Ruiz et al. (2005) identified tph1 recessive gene in both T589 and LG15 lines. Demurin et al. (2004) identified *tph2* gene in lines T2100 and LG-17. Using these lines, the first sunflower hybrids with higher tocopherol composition were developed. The hybrid Krasnodarskiy 917 has 50% of β-tocopherol content. Another hybrid Oxy has 60% of γ- and 40% of β-tocopherol content with high oleic acid (Demurin 2012).

Table 1 Variations in gene action for high oleic acid in sunflower

Dominant/recessive/ intermediary	Major gene + modifiers	References
Partially dominant	1 major gene	Fick (1984)
Dominant	1 major gene + modifiers	Urie (1985)
Intermediate	1 major gene + modifiers	Miller et al. (1987)
Dominant	1 major gene	Schmidt et al. (1989)
Dominant	3 major genes additive + modifiers	Martinez et al. (1989)
Dominant/partially dominant	No conclusion	Nikolova et al. (1991)
Dominant, sometimes recessive	3 hypotheses as to increasing gene numbers	Demurin and Skoric (1996)
Dominant, sometimes recessive	1 major gene + modifiers not clear	Dehmer and Friedt (1998)
Partially dominant	2 interacting major genes	Martinez et al. (1999)
Dominant	Not addressed	Vares et al. (2000)
Dominant	Major QTL (85% EV)	Perezvich et al. (2000)
Complex, some maybe dominant	Five major genes + modifiers	Velasco et al. (2000)
Dominant	1 M locus = Δ 12-RFLP	Lacombe et al. (2001)
Dominant	One or more genes with modifier effect	Berville (2010)
Dominant	At least three loci	Ferfuia and Vannozzi (2015)
Dominant	Three QTLs	Premnath et al. (2016)

5 Genetic Diversity Analysis for Oil Quality Traits in Sunflower

Cultivated accessions of the sunflower have useful genes that encode various phenotypic traits that could be used to speculate on the origin of sunflower oil. These genes may later be employed in breeding initiatives to produce sunflower oil of higher quality. Sunflower oil has the potential for enhanced nutritional properties. Breeders have a variety of approaches for increasing oil content, including using genetic variation from several sunflower germplasms and wild populations. The traits can be introduced into grown sunflowers by gathering, researching, and choosing the best populations (Seiler 2007). Although wild accessions of sunflowers have less seed oil than cultivated sunflowers, backcrossing can increase the oil content to a desirable level (Seiler and Gulya 2010).

5.1 Phenotype Based Diversity Analysis for Quality Traits in Sunflower

The major objective of sunflower breeding is to develop hybrids or varieties with high yield, oil content, and oil quality along with disease resistance. To achieve these

objectives, the breeder needs to select genetically diverse parents. Therefore, knowledge of the existing genetic diversity in the germplasm is important to plan for recombination breeding.

Ahmadian et al. (2019) assessed the genetic diversity of 107 cultivated accessions of sunflower. A significant negative relationship between oleic acid, stearic acid, and saturated fatty acids (SFA) with oil content was observed. The correlation coefficient of the ratio between unsaturated fatty acids to saturated fatty acids (UFA/SFA ratio) with oil content was positive and significant. In the PCA analysis, four major principal components (PCs) were identified. These PCs account for 87.19% of the total variation. The accessions were grouped into seven distinct clusters, and the accessions in clusters 4 and 7 contained high UFA and low SFA values.

Dudhe et al. (2019) evaluated 2149 sunflower germplasm accessions. Significant differences among genotypes for almost all characters were observed. The first PC accounted for 29.20% of the total variation in the population with more contribution from oil content. The second PC contributed 57.6% to days to maturity and 50% flowering contributed maximum. DUS characters such as pigmentation of seedlings, leaf petiole, disc, and stem can be considered as morphological markers to differentiate the germplasm. Ray floret coloration, plant branching, type of branching, and pollen color characters can help the breeder to identify the specific germplasm. The identified trait-specific accessions will help in the effective utilization of promising accessions in the breeding.

5.2 Marker Based Diversity Analysis in Sunflower

Molecular markers play a major role in identifying variation in genomic DNA sequences. Polymorphism at the molecular level in different species has played a major role in the analysis of genetic diversity, identification of phylogenetic relationships, and also transfer of target traits to elite germplasm. Different molecular markers have been used in mapping genes leading to the development of sunflower linkage maps resulting in studies revealing genetic diversity in the genus *Helianthus* (Mwangi et al. 2019).

Mandel et al. (2011) studied an association panel of sunflower. It consists of 433 cultivated and 24 wild accessions. A set of 34 expressed sequence tag (EST)-simple sequence repeats (SSRs) were used. Gene diversity was 0.47 and 0.70 in cultivated and wild accessions, respectively. The results showed that cultivated accessions had less genetic diversity than wild accessions. Further wild accessions were grouped in four clusters. The cultivated accessions were grouped into two clusters. The maximum likelihood method indicated that the diversity in cultivated accessions was sourced from two wild sunflower populations. These are from the east-central USA. It is the same region where sunflower domestication is presumed to have occurred. A nested subset of accessions captured more allelic diversity in the cultivated accessions. A core set of 288 accessions captured about 90% of alleles of full set. Another core set of just 12 captured about 50% of alleles in cultivated accessions.

Filippi et al. (2015) studied the association mapping population. This population consisted of various accessions. It had 137 genotypes of National Institute of

Agricultural Technology, Argentina. A set of 33 open-pollinated and composite populations were also included. Two types of markers, viz., 42 SSR markers and single-nucleotide polymorphism (SNP) markers, were used. Clustering was compared for the two marker types. The population structure was dominated by the maintainer/restorer trait.

The molecular diversity in a set of 114 cultivated sunflower populations was studied by single-nucleotide polymorphism genotyping (Mangin et al. 2017a). The mean allele number varied from 1.07 to 1.90. Intrapopulation variability was slightly reduced according to the number of multiplications since entry, but some populations were probably largely homozygous when received. A principal component analysis was used to study interpopulation variability. The first three axes accounted for 17% of total intrapopulation variability. The first axis was significantly correlated with seed oil content, more closely than just the distinction between oil and confectionary types. The second axis was related to the presence or absence of restorer genes, and the third axis was to flowering date and possibly to adaptation to different climates. Propositions are made to improve sunflower population maintenance procedures to keep maximum genetic variability for future breeding (Mangin et al. 2017a).

Chernova et al. (2021) performed high-throughput lipidomic profiling in sunflower by genotyping 601 inbred sunflower lines and analyzed their lipid and fatty acid composition. The genome-wide association analysis based on genotypes for 15,483 SNPs and the concentrations of 23 fatty acids, including minor fatty acids, revealed significant genetic associations for 11 of them. The identified genomic regions included the loci involved in rare fatty acid variation on chromosomes 3 and 14, explaining up to 34.5% of the total variation of docosanoic acid (22:0) in sunflower oil.

6 Mutation Breeding for Altered Oil Quality in Sunflower

A substantial change in the quality of sunflower was created through mutation breeding. In contrast with other oil crops, modifications in oil quality are considerably more in sunflower. This was possible through many factors. These factors are induced mutations with both physical and chemical means, precise phenotyping techniques to fix mutants, nondestructive analyses, etc. Mutants have various levels of saturated fatty acids, oleic acid content, β-, γ- and δ-tocopherols. It helped to create more oil profiles in sunflower. For example, sunflower oil contains the above 90% oleic acid. It is higher than any other vegetable oils in the world market (Cjevic et al. 2014). Fatty acid and tocopherol content had higher stability over environments. These traits are controlled by a few genes. Hence, it can easily be utilized to develop hybrids with different oil quality.

6.1 Mutation Breeding for Altered Fatty Acid Composition in Sunflower

The outcome of mutation breeding for modified fatty acid content composition is the changes in their proportion. Mutation with DMS resulted in high oleic

acid content (>80%). EMS and sodium azide (NaN_3) mutagens provided increased stearic acid content (over 25%). Likewise higher proportion of palmitic acid content was obtained with the usage of physical mutagens (X-rays and γ-rays). The high oleic acid content donor was developed by Soldatov (1976). He used DMS to obtain mutants from VNIMK 8931 variety. His studies resulted in the identification of the high oleic (80–90%) Pervenet variety. Pervenet has been widely used as a high oleic donor in sunflower improvement programs.

6.2 Mutation Breeding for Altered Tocopherol Content in Sunflower

Kurnik (1966) identified a mutant with modified tocopherol composition. He reported the absence of α-tocopherol and presence of β- and γ-tocopherols in two samples of Peredovik variety. Demurin (1986) reported a spontaneous mutation with 50% increase in β-tocopherol content. It was identified with the use of half-seed technique and TLC (Popov and Aspiotis 1991). The line LG15 with higher β-tocopherol content was identified in this material. Another spontaneous mutant LG17 was identified in the USA with modified tocopherol. It had 5% of α- and 95% of γ-tocopherol (Popov and Demurin 1987).

7 Molecular Mapping of Fatty Acid Related Genes in Sunflower

Genotype of female parent influences the oleic acid content (OAC) (Ferfuia et al. 2015). Increased OAC was obtained through silencing the *FAD2-1* gene. The gene encoding *FAD2* enzyme was obtained through mutation (Schuppert et al. 2006). It controls the pathway of oleic-linoleic conversion. The Pervenet variety is the main donor for higher OAC in sunflower improvement programs (Allman-Farinelli et al. 2005). The OAC was controlled by one dominant to many genes (Urie 1985; Lacombe et al. 2004; Joksimovic et al. 2006; Berville 2010; Ferfuia and Vannozzi 2015; Premnath et al. 2016; Dimitrijevic et al. 2017).

Various markers were used to map *Ol* mutation in sunflower. Dehmer and Friedt (1998) mapped *Ol1* gene with random amplified polymorphic DNA (RAPD) markers F15-690 and AC10-765 with flanking distance of 7.0 and 7.2 cM. Perez-Vich et al. (2002) and Schuppert et al. (2006) mapped *Ol1* on linkage group (LG)14. Perez-Vich et al. (2002) mapped major quantitative trait locus (QTL) for OAC with 84.5% as R^2. Schuppert et al. (2006) reported 49 SNPs and 5 insertion/deletions (INDELs) to identify Ol mutation. Tightly linked codominant and dominant SSR markers were reported by Lacombe et al. (2009). Premnath et al. (2016) reported three quantitative trait loci (QTLs) at LG8, LG9, and LG14. The QTLs on LG14 and LG8 had more than 60% as R^2 for OAC. Dimitrijevic et al. (2017) reported that marker F4-R1 was effective for selection of high OAC.

8 Marker-Assisted Breeding for Oil Quality Traits in Sunflower

Molecular markers were efficiently used in sunflower improvement (Dimitrijevic and Horn 2018). These markers are highly polymorphic and mostly codominant. It has linkage with the trait of interest. These markers can be used at all growth stages and phenotypically neutral. QTL analysis helps to identify potential markers for yield and component traits (Abdi et al. 2012), high OAC (Schuppert et al. 2006; Dimitrijevic et al. 2017), β-tocopherol content (Vera-Ruiz et al. 2006), and γ-tocopherol content (Moreno et al. 2006).

8.1 MAS for High Oleic Acid Content in Sunflower

Breeding for high OAC is an essential objective in sunflower. Normally sunflower has 18–25% OAC (Rauf et al. 2017). As OAC has beneficial effect on human health, this trait was incorporated into hybrids. The dominant mutation in Pervenet variety increases OAC to >89%. Commercial varieties with high OAC have a premium in price (Rauf et al. 2017). At present, these varieties occupy up to 4% of the total sunflower oil production. Phenotypic selection for high OAC is expensive and slow. It involves the laborious gas chromatography protocols. Molecular markers facilitate the selection in early generations. Markers like NI-3F/N2-IR help to identify genotypes in segregating material (Rauf et al. 2020).

8.2 MAS for Tocopherol Content in Sunflower

Tocopherols are highly abundant in sunflower oil (Rauf et al. 2017). This fat-soluble tocopherol has vitamin E activity. These tocopherols protect cells and oil from oxidative damage (Rauf et al. 2017). Sunflower seed is predominantly comprised of α-tocopherol. Substitution of α- with γ-tocopherol helps to improve the shelf life of the oil (Moreno et al. 2012). The selection for tocopherol content is expensive and laborious. Moreover, this trait has complex genetics due to interaction of loci for tocopherol derivatives (Moreno et al. 2012).

Molecular markers help to select plants with desired composition of fatty acids in oil with low cost. *Tph1* gene controls the β-tocopherol content. A QTL with flanking markers ORS-1093, ORS-222, and ORS-598 on LG1 is associated with *Tph1*. Allelic pattern of ORS-716 was useful to identify the low β-tocopherol and high β-tocopherol genotypes (Vera-Ruiz et al. 2006). Moreno et al. (2006) developed inbred lines with high γ-tocopherol content (85%). They also reported that *Tph2* controls the levels of γ-tocopherol. The flanking region of markers ORS-312 and ORS-599 on LG8 had association with γ-tocopherol content (Moreno et al. 2006).

9 Association Mapping Studies in Sunflower

Association mapping (also known as linkage disequilibrium (LD) mapping) has emerged as an alternative to QTL mapping for investigating the genetic basis of quantitative traits. Because it involves the analysis of a diverse collection of more or less unrelated individuals, association mapping allows for the simultaneous evaluation of the effects of multiple haplotypes across diverse genetic backgrounds. Moreover, because association populations typically capture numerous generations of historical recombination, this approach provides much higher resolution than is possible with a family-based mapping population. Nambeesan et al. (2015) reported the results of a detailed analysis of variation in branching in an association mapping population that captures nearly 90% of the allelic diversity present within the cultivated sunflower gene pool (Mandel et al. 2011, 2013).

Two approaches have been followed in association mapping. They are genome-wide association studies (GWAS) and candidate gene approaches. High-throughput marker systems nowadays give full genome coverage. It helps studies like GWAS, QTL Seq, and genomic selection possible. Fusari et al. (2008) reported that the linkage disequilibrium in sunflower rapidly decays. Hence, association studies may help to detect QTLs. However, assessment of population structure should be checked to avoid false positives. So far, Mandel et al. (2013) alone reported one association mapping. All other reports were on candidate gene approach (Fusari et al. 2012; Cadic et al. 2013; Talukder et al. 2014; Nambeesan et al. 2015; McAssey et al. 2016).

Mandel et al. (2011) studied an association population for GWAS with SNP markers. Traits, viz., flowering time, branching, and heterotic groups were studied. LD showed significant association between marker and trait. They reported that selection for disease resistance and domestication were the factors for the genome-wide variations. Nambeesan et al. (2015) carried out an association mapping study based on candidate genes for branching. A total of 39 genes for branching were detected. Up to eight of the highest BLAST hit for each gene were included in the analyses due to the recent triplication of the sunflower genome (Badouin et al. 2017). SNPs associated with 13 candidate genes for branching were identified (Nambeesan et al. 2015). Most of these were found on LG10. Earlier QTL studies also mapped the B-locus for branching on LG10 (Tang et al. 2006; Bachlava et al. 2009). McAssey et al. (2016) reported a SNP linked with flowering time QTL.

Both association and linkage mapping were used to identify QTL for flowering time (Cadic et al. 2013). Associations with flowering time could be demonstrated for 11 regions distributed over 10 LGs. In addition, QTLs for flowering time were detected on 11 LGs in a RIL population by linkage mapping. This large number of QTL is consistent with the polygenic pattern of inheritance of flowering time reported earlier (Leon et al. 2000). SNPs detected by association mapping were then investigated about positional overlaps with QTL identified in the RIL population. The remaining eight regions contained five candidate genes linked with

flowering time in other species that showed SNPs in sunflower, and one of the genes was the gibberellin receptor GID1B (Cadic et al. 2013). Thirty genes, including this gene, had before been investigated as candidate genes for flowering time concerning domestication and improvement in sunflower (Blackman et al. 2011). Kiani et al. (2009) reported a major QTL through linkage mapping. However, it was not found by association mapping (Cadic et al. 2013). It may be due to the lower proportion of alleles in the association population.

10 Genomic Selection for Oil Quality Traits in Sunflower

Individuals were selected using genomic breeding values in genomic selection (GS). The concept of GS is to use genome-wide molecular information to identify QTLs. Many markers were developed by genotyping-by-sequencing. New 25 K SNP genotyping arrays (Livaja et al. 2016) and sequencing of parental lines (Mangin et al. 2017b) were also helped to develop more markers. The hybrid performance was predicted with 572 AFLP markers (Reif et al. 2013). GS helps to accurately predict the hybrid performances when the parents were closely related (Reif et al. 2013). However, prediction based on GCA was not improved by GS.

Prediction accuracy of GS and GCA modeling was compared with 36 CMS lines and 36 restorer lines (Mangin et al. 2017b). Multi-environmental field trials were carried out with 452 hybrids for oil content. All parents were sequenced to obtain SNP markers (Mangin et al. 2017b). Mangin et al. (2017b) reported that GS had more efficiency than GCA modeling.

11 Recent Concepts and Strategies Developed in Sunflower

Advances in molecular research and the availability of vast amounts of genomic data offer opportunities to overcome breeding challenges and ensure the advancement of sunflower, as an important source of edible oil. In recent years, dense genetic maps, largely based on SNPs, have become available; hence, markers can be found for major genes or QTL. The availability of this reference genome and companion resources will not only strengthen interest in the sunflower as a model for ecological and evolutionary studies but will also accelerate breeding programs (Badouin et al. 2017).

11.1 Whole-Genome Sequence Information of Sunflower

The domesticated sunflower is an important oilseed crop. It has promise for climate change adaptation. It can maintain stable yields across a wide variety of environmental conditions. Even more resilience is obtainable via the mining of resistance alleles from crossable wild accessions. Badouin et al. (2017) provided a high-quality reference for the sunflower genome (3.6 gigabases), together with enormous

transcriptomic data from vegetative and floral organs. The genome mostly consists of highly similar, related sequences and required single-molecule real time sequencing technologies for successful assembly. Genome analyses enabled the reconstruction of the evolutionary history of the Asterids. Further, it established the existence of a whole-genome triplication at the base of the Asterids II clade and sunflower specific whole-genome duplication around 29 million years ago (Barker et al. 2016). An integrative approach combining quantitative genetics, expression, and diversity data permitted the development of comprehensive gene networks for flowering time and oil metabolism. It also revealed new candidate genes in these networks. The sunflower genome represents a cornerstone for future research programs. It may help to exploit genetic diversity to improve biotic and abiotic stress resistance and oil production.

11.2 Integrated Omics Approaches in Sunflower

Omics technologies offer novel possibilities for deciphering the complex pathways and molecular profiling through the level of systems biology and can provide important answers that can be utilized for more efficient breeding of sunflower. The development of sunflower as a significant source of edible oil can be ensured due to developments in genetic research and the accessibility of enormous amounts of genomic data. A thorough description of genotypic and phenotypic variation is crucial for gaining a greater understanding of research that has been done into the variations of critical features that can be used to enhance sunflower genetics. Using an integrated strategy involving proteomics, metabolomics, phenomics, transcriptomics, epigenomics, and genomics in an integrated manner termed as systems biology provides fresh opportunities to understand intricate networks and molecules profiling for significant and intricate agronomic features and deliver solutions that can improve sunflower (Jockovic et al. 2021).

11.3 Initiation of Genome Editing in Sunflower

Over the recent years, clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) systems have been widely leveraged to edit the genomic DNA in many organisms. The prevailing CRISPR-based genome editing systems, for instance, CRISPR/Cas9, CRISPR/Cpf1, base editing system, and prime editing system, have brought promises to highly effective improvement of crop genetics with elite traits since they are born (Zhu et al. 2020). Compared to the zinc finger nuclease system (ZFN system) and transcription activator-like effector nuclease system (TALEN system), CRISPR-based genome editing technology has unparalleled advantages in easy manipulation. Indeed, theoretically, we can mutate almost every gene of interest only by rationally designing the target sequence and expressing the editing systems in targeted crops. Accounting for the unprecedented merits and the easy access to genome sequences of ever-increasing plant species,

CRISPR-based genome editing systems have broad applications in fundamental biological research and crop breeding. There are very few reports on the applications of CRISPR-based genome editing technology in sunflower, mainly due to its complicated genomes and recalcitrance to transformation (He et al. 2021).

12 Conclusion and Future Prospects

Sunflower genome database allows access to transcriptome data. These were effectively used to address flowering time and oil metabolism (Badouin et al. 2017). GWAS and GS with large amounts of markers are useful to address complex traits in sunflower. Considering the ever-increasing newly developed and diverse CRISPR/Cas systems and their derivative editors, with the continuous development of CRISPR/Cas reagent delivery methods such as carbon nanomaterial mediated genetic elements delivery (Demirer et al. 2019), genome editing in sunflower would flourish soon. The new genomic tools may help to understand the genetic variation in sunflower. These techniques may help to evolve a variety with combination of several quality traits. This will enable tailoring speciality oils for food and nonfood industry. These developments help to achieve a promising future for sunflower in the world market.

References

Abdi N, Darvishzadeh R, Jafari M et al (2012) Genetic analysis and QTL mapping of agro-morphological traits in sunflower (*Helianthus annuus* L) under two contrasting water treatment conditions. Plant Omics 5(2):149–158

Adeleke BS, Babalola OO (2020) Oilseed crop sunflower (*Helianthus annuus*) as a source of food: nutritional and health benefits. Food Sci Nutr 8:4666–4684. https://doi.org/10.1002/fsn3.1783

Ahmadian S, Enferadi ST, Alemzadeh A (2019) Assessment of genetic diversity of cultivated sunflower in terms of oil content, fatty acid compositions and seed traits. Helia 9. https://doi.org/10.1515/helia-2019-0009

Allman-Farinelli MA, Gomes K, Favaloro EJ, Petocz P (2005) A diet rich in high-oleic-acid sunflower oil favorably alters low-density lipoprotein cholesterol, triglycerides, and factor VII coagulant activity. J Am Diet Assoc 105:1071–1079. https://doi.org/10.1016/j.jada.2005.04.008

Bachlava E, Tang S, Pizarro G, Knapp SJ (2009) Pleiotropy of the branching locus (B) masks and unlinked quantitative trait loci affecting seed traits in sunflower. Theor Appl Genet 120:829–842. https://doi.org/10.1007/s00122-009-1212-1

Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L et al (2017) The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. Nature 546:148–152. https://doi.org/10.1038/nature22380

Baldini M, Vannozzi GP (1998) Agronomic and physiological assessment of genotypic variation for drought tolerance in sunflower genotypes obtained from a cross between *H. annuus* and *H. argophyllus*. Agric Med 128:232–240

Barker MS, Li Z, Kidder TI et al (2016) Most Compositae (Asteraceae) are descendants of apaleohexaploid and all share a paleotetraploid ancestor with the Calyceraceae. Am J Bot 103:1203–1121

Berville A (2010) Oil composition variations. In: Hu J, Seiler G, Kole C (eds) Genetics, genomics and breeding of sunflower. Science Publisher, Boca Raton, pp 253–277

Blackman BK, Rasmussen DA, Strasburg JL et al (2011) Contribution of flowering time genes to sunflower domestication and improvement. Genetics 187:271–287. https://doi.org/10.1534/genetics.110.121327

Bowsher AW, Milton EF, Donovan LA (2016) Comparison of desert-adapted *Helianthus niveus* (Benth.) Brandegee ssp. *tephrodes* (A. Gray) Heiser to cultivated *H. annuus* L. for putative drought avoidance traits at two ontogenetic stages. Helia 39(64):1–19. https://doi.org/10.1515/helia-2016-0003

Cadic E, Coque M, Vear F, Grezes-Besset B, Puaquet J, Piquemal J et al (2013) Combined linkage and association mapping of flowering time in sunflower (*Helianthus annuus* L.). Theor Appl Genet 126:1337–1356. https://doi.org/10.1007/s00122-013-2056-2

Cheevarungnapakul K, Khaksar G, Panpetch P, Boonjing P, Sirikantaramas S (2019) Identification and functional characterization of genes involved in the biosynthesis of caffeoylquinic acids in sunflower (*Helianthus annuus* L.). Front Plant Sci 10:968. https://doi.org/10.3389/fpls.2019.00968

Chernova AI, Gubaev RF, Singh A et al (2021) Genotyping and lipid profiling of 601 cultivated sunflower lines reveals novel genetic determinants of oil fatty acid content. BMC Genomics 22:505. https://doi.org/10.1186/s12864-021-07768-y

Cjevic S, Miladinovic D, Jocic S (2014) Mutation breeding for changed oil quality in sunflower. In: Tomlekova NB, Kozgar MI, Wani MR (eds) Mutagenesis: exploring genetic diversity of crops. Wageningen Academic Publishers, Wageningen. https://doi.org/10.3920/978-90-8686-796-7_4

Dehmer KJ, Friedt W (1998) Development of molecular markers for high oleic acid content in sunflower (*Helianthus annuus* L.). Ind Crop Prod 7:311–315

Delplanque B (2000) The nutritional value of sunflower oils: linoleic sunflower seeds and seeds with high oleic content. OCL 7(6):467–472

Demirer GS, Zhang H, Matos JL et al (2019) High aspect ratio nano materials enable delivery of functional genetic material without DNA integration in mature plants. Nat Nanotechnol 14:456–464

Demurin Y (1986) Genetic analysis of tocopherol composition in sunflower seeds. Sci Techn Bull VIR 165:49–51

Demurin Y (1993) Genetic variability of tocopherol composition in sunflower seeds. Helia 16(18):59–62

Demurin Y (2003) Up-to-date results on biochemical genetics of sunflower in VNIIMK. Helia 26(38):137–141

Demurin Y (2012) Tocopherol mutations in sunflower. In: ISA-ASAGIR (ed) Program and Abstract of the 18th international sunflower conference, Mar del Plata/Balcarce, 27 Feb – 1 Mar 2012, p 51

Demurin Y, Skoric D (1996) Unstable expression of *Ol* gene for high oleic acid content in sunflower seeds. In: Proceedings of the 14th international sunflower conference, Beijing/Shenyang, 12–20 June 1996, pp 145–150

Demurin Y, Efimenko SG, Peretyagina TM (2004) Genetic identification of tocopherol mutations in sunflower. Helia 27(40):113–116

Dimitrijevic A, Horn R (2018) Sunflower hybrid breeding: from markers to genomic selection. Front Plant Sci 8:2238. https://doi.org/10.3389/fpls.2017.02238

Dimitrijevic A, Imerovski I, Miladinovic D, Cvejic S, Jocic S, Zeremski T, Sakac Z (2017) Oleic acid variation and marker-assisted detection of Pervenets mutation in high- and low-oleic sunflower cross. Crop Breed Appl Biotechnol 17:229–235

Dudhe MY, Mulpuri S, Meena HP et al (2019) Genetic variability, diversity and identification of trait-specific accessions from the conserved sunflower germplasm for exploitation in the breeding programme. Agric Res 9:9–22

Ferfuia C, Vannozzi GP (2015) Maternal effect on seed fatty acid composition in a reciprocal cross of high oleic sunflower (*Helianthus annuus* L.). Euphytica 205:325–336. https://doi.org/10.1007/s10681-015-1378-3

Ferfuia C, Turi M, Vannozzi GP (2015) Variability of seed fatty acid composition to growing degree-days in high oleic acid sunflower genotypes. Helia 38:61–78

Fick GN (1984) Inheritance of high oleic acid in the seed oil of sunflower. In: Proceedings of the sunflower research workshop. National Sunflower Association, Bismarck, p 9

Filippi CV, Aguirre N, Rivas JG et al (2015) Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers. BMC Plant Biol 15:52. https://doi.org/10.1186/s12870-014-0360-x

Fusari CM, Lia VV, Hopp HE, Heinz RA, Paniego NB (2008) Identification of single nucleotide polymorphisms and analysis of linkage disequilibrium in sunflower elite inbred lines using the candidate gene approach. BMC Plant Biol 8:7. https://doi.org/10.1186/147-2229.8-7

Fusari CM, Di Rienzo JA, Troglia C, Nishinakamasu V, Moreno MV, Maringolo C et al (2012) Association mapping in sunflower for *Sclerotinia* head rot resistance. BMC Plant Biol 12:93. https://doi.org/10.1186/1471-2229-12-93

Guo S, Ge Y, Jom KN (2017) A review of phytochemistry, metabolite changes, and medicinal uses of the common sunflower seed and sprouts (*Helianthus annuus* L.). Chem Central J 11:95. https://doi.org/10.1186/s13065-017-0328-7

Hajjar R, Hodgkin T (2007) The use of wild relatives in crop improvement: a survey of developments over the last 20 years. Euphytica 156:1–13

He J, Zhang K, Tang M et al (2021) CRISPR-based genome editing technology and its applications in oil crops. Oil Crop Sci 6:105–113

Heiser CB, Smith DM, Clevenger SB, Martin WC (1969) The north American sunflowers (*helianthus*). Memoirs Torr Bot Club 22:1–218

Islam RT, Hossain MM, Majumder K, Tipu AH (2016) *In vitro* phytochemical investigation of *Helianthus annuus* seeds. Bang Pharmaceut J 19(1):100–105. https://doi.org/10.3329/bpj.v19i1.29245

Jocic S, Miladinovic D, Jocic Y (2015) Breeding and genetics of sunflower. In: Enrique Martínez-Force E, Dunford NT, Salas JJ (eds) Sunflower: chemistry, production, processing, and utilization. Academic and AOCS Press, Urbana, pp 1–25

Jockovic M, Jocic S, Cvejic S et al (2021) Genetic improvement in sunflower breeding – integrated omics approach. Plan Theory 10:1150

Joksimovic J, Atlagic J, Marinkovic R, Jovanovic D (2006) Genetic control of oleic and linoleic acid contents in sunflower. Helia 29:33–40. https://doi.org/10.2298/hel0644033j

Karangwa E, Zhang X, Murekatete N et al (2015) Effect of substrate type on sensory characteristics and antioxidant capacity of sunflower Maillard reaction products. Eur Food Res Technol 240(5): 939–960

Kiani PS, Maury P, Nouri L, Ykhef N, Grieu P, Sarrafi A (2009) QTL analysis of yield-related traits in sunflower under different water treatments. Plant Breed 128:363–373. https://doi.org/10.1111/j.1439-0523.2009.01628.x

Kurnik E (1966) Zuchtung von Sonnenblumen (*Helianthus annuus*) unter Berucksichtigung von Olmenge undolqualitat. Qualitas Plantarum et Materiae Vegetabilies 13(1–4):157–170

Lacombe S, Kaan F, Leger S, Berville A (2001) An oleate desaturase and a suppressor loci direct high oleic acid content of sunflower (*Helianthus annuus* L.) oil in the Pervenets mutant. Life Sci (Paris) 324:839–845

Lacombe S, Kaan F, Griveau Y, Berville A (2004) The Pervenets high oleicmutation: methodological studies. Helia 27:41–54. https://doi.org/10.2298/hel04400411

Lacombe S, Souyris I, Berville AJ (2009) An insertion of oleate desaturasehomologous sequence silences *via* siRNA the functional gene leading to high oleic acid content in sunflower seed oil. Mol Genet Genomics 281:43–54

Leon AJ, Andrade FH, Lee M (2000) Genetic mapping of factors affecting quantitative variation for flowering in sunflower. Crop Sci 40:404–407

Livaja M, Unterseer S, Erath W, Lehermeier C, Wieseke R, Plieske J et al (2016) Diversity analysis and genomic prediction of *Sclerotinia* resistance in sunflower using a new 25 K SNP genotyping array. Theor Appl Genet 129:317–329

Ludwig F, Rosenthal DM, Johnston JA et al (2004) Selection on leaf ecophysiological traits in a desert hybrid *Helianthus* species and early-generation hybrids. Evolution 58:2682–2692

Mandel JR, Dechaine JM, Marek LF, Burke JM (2011) Genetic diversity and population structure in cultivated sunflower and a comparison to its wild progenitor, *Helianthus annuus* L. Theor Appl Genet 123:693–704

Mandel JR, Nambeesan S, Bowers JE, Marek L, Ebert D, Rieseberg LH et al (2013) Association mapping and the genomic consequences of selection in sunflower. PLoS Genet 9:e1003378. https://doi.org/10.1371/journal.pgen.1003378

Mangin B, Pouilly N, Boniface MC, Langdale NB, Vincourt P, Vear F et al (2017a) Molecular diversity of sunflower populations maintained as genetic resources is affected by multiplication processes and breeding for major traits. Theor Appl Genet 130:1099–1112. https://doi.org/10.1007/s00122-017-2872-x

Mangin B, Bonnafous F, Blanchet N, Boniface MC, Bret-Mestries E, Carrere S et al (2017b) Genomic prediction of sunflower hybrids oil content. Front Plant Sci 8:1633. https://doi.org/10.3389/fpls.2017.01633

Martinez JF, Jimenez A, Dominguez J, Magarcia J, Garces R, Mancho M (1989) Genetic analysis of the high oleic acid content in cultivated sunflower (*Helianthus annuus* L.). Euphytica 41:39–51

Martinez HF, Baldini M, Olivieri AM (1999) Inheritance of oleic acid content in sunflower oil. J Genet Breed 53:99–103

McAssey E, Corbi J, Burke JM (2016) Range-wide phenotypic and genetic differentiation in wild sunflower. BMC Plant Biol 16:249. https://doi.org/10.1186/s12870-016-0937-7

Moreno MJG, Ruiz EMV, Martinez JMF, Velasco L (2006) Genetic and Molecular Analysis of High Gamma-Tocopherol Content in Sunflower. Crop Sci 46(5):2015.

Miller JF (1995) Inheritance of salt tolerance in sunflower. Helia 18:9–16

Miller JF, Zimmerman DC, Vick BA (1987) Genetic control of high oleic acid content in sunflower oil. Crop Sci 27(5):923–926

Moreno MJG, Martínez JMF, Velasco L et al (2012) Genetic basis of unstable expression of high gamma-tocopherol content in sunflower seeds. BMC Plant Biol 12(1):71. http://www.biomedcentral.com/1471-2229/12/71

Mwangi EW, Marzougui S, Sung JS et al (2019) Assessment of genetic diversity and population structure on kenyan sunflower (*Helianthus annus* L.) breeding lines by SSR markers. Kor J Plant Res 32(3):244–253

Nambeesan SU, Mandel JR, Bowers JE et al (2015) Association mapping in sunflower (*Helianthus annuus* L.) reveals independent control of apical *vs*. basal branching. BMC Plant Biol 15:84. https://doi.org/10.1186/s12870-015-0458-9

Nikolova V, Ivanov P, Petakov D (1991) Inheritance of sunflower oil quality following crossing of lines with different fatty acid composition. Genet Breed 24:391–397

Oliveira VS, Oliveira RL, Goes R, Silva TM, Silva LF, Freitas LS, Pereirad ES, Bezerra LR (2019) Physicochemical composition, fatty acid profile and sensory attributes of the meat of young Nellore bulls fed sunflower cake from the biodiesel industry. Livestock Sci 227:97–104

Packer L, Jevic UC (2002) Vitamin E: an introduction. In: Packer L, Traber MG, Kraemer K, Frei B (eds) The antioxidants vitamins C and E. AOCS Press, Champaign, pp 133–151

Perezvich B, Martinez JMF, Knapp SJ, Berry ST(2000) Molecular markers associated with sunflower oleic and stearic acid concentrations in a high stearic x high oleic cross. In: Proceedings of the XV sunflower international conference, Toulouse, pp 55–60

Perez-Vich B, Garces R, Martinez JMF (2002) Inheritance of high palmitic acid and its relationship with high oleic acid content in the sunflower mutant CAS 12. Plant Breed 121:49–56

Popov PS, Aspiotis EH (1991) Biochemical methods for breeding material estimation. In: Biology, breeding and growing of sunflower. Agropromizdat, Moscow, pp 77–80

Popov PS, Dyakov AB, Borodulina AA, Demurin Y (1988) Genetic analysis of tocopherol and fatty acid composition in sunflower seeds. Genetics 24(3):518–527

Popov PS, Demurin Ya N (1987) Mutation variability and inheritance of tocopherol composition in sunflower seeds of VIRcollection (in Russian). Breeding and genetics of technical crops.VIR, Leningrad 113:30–33

Premnath A, Narayana M, Ramakrishnan C, Kuppusamy S, Chockalingam V (2016) Mapping quantitative trait loci controlling oil content, oleic acid and linoleic acid content in sunflower (*Helianthus annuus* L.). Mol B 36:106. https://doi.org/10.1007/s11032-016-0527-2

Rauf S, Jamil N, Tariq SA et al (2017) Progress in modification of sunflower oil to expand its industrial value. J Sci Food Agric 97:1997–2006

Rauf S, Warburton M, Naeem A, Kainat W (2020) Validated markers for sunflower (*Helianthus annuus* L.) breeding. OCL 27:47. https://doi.org/10.1051/ocl/2020042

Reif JC, Zhao Y, Wurschum T, Gowda M, Hahn V (2013) Genomic prediction of sunflower hybrid performance. Plant Breed 132:107–114. https://doi.org/10.1111/pbr.12007

Schmidt L, Marquard R, Friedt W (1989) Status and prospects of breeding high oleic acid sunflowers for Central Europe. Fat Sci Technol 91:346–349

Schuppert GF, Tang S, Slabaugh MB, Knapp SJ (2006) The sunflower high-oleic mutant *Ol* carries variable tandem repeats of *FAD2-1*, a seed specific oleoyl-phosphatidyl choline desaturase. Mol Breed 17:214–256

Seiler GJ (2007) Wild annual *Helianthus anomalus* and *H. deserticola* forimproving oil content and quality in sunflower. Ind Crop Prod 25:95–100

Seiler GJ, Gulya TJ (2010) Evaluation of *Helianthus porteri* achenes for oil concentration and fatty acid composition. Botanical Society of America, Providence, 31 July 31–4 Aug 2010, p 82

Seiler GJ, Qi LL, Marek LF (2017) Utilization of sunflower crop wild relatives for cultivated sunflower improvement. Crop Sci 57:1083–1101

Skoric D, Jocic S, Lecic N, Sakac Z (2008) Genetic possibilities for altering sunflower oil quality to obtain novel oils. Can J Physiol Pharmacol 86(4):215–221

Soldatov KI (1976) Chemical mutagenesis in sunflower breeding. In: Proceedings of the 7th International Sunflower Association, Vlaardingen, pp 352–357

Stebbins JC, Winchell CJ, Constable JVH (2013) *Helianthus winteri* (Asteraceae), a new perennial species from the southern Sierra Nevada foothills, California. Aliso 31:19–24

Talukder ZI, Gong L, Hulke BS, Pegadaraju V, Song Q, Schultz Q et al (2014) A high-density SNP map of sunflower derived from RAD-sequencing facilitating fine-mapping of the rust resistance gene R12. PLoS One 9:e98628. https://doi.org/10.1371/journal.pone.0098628

Tang S, Leon A, Bridges WC, Knapp SJ (2006) Quantitative trait loci for genetically correlated seed traits are tightly linked to branching and pericarp pigment loci in sunflower. Crop Sci 46:721–734. https://doi.org/10.2135/cropsci2005.0006-7

Terzic S, Boniface MC, Marek L et al (2020) Gene banks for wild and cultivated sunflower genetic resources. OCL 27(9):14. https://doi.org/10.1051/ocl/2020004

Urie AL (1985) Inheritance of high oleic acid in sunflower. Crop Sci 25(6):986–989

Vares D, Cleomene J, Lacombe S, Griveau Y, Berville A, Kaan F (2000) Triacyl glyceride composition in F_1 seed using factorial and diallel crosses between sunflower lines. In: Proceedings of the 15th international sunflower conference, Toulouse, pp 19–24

Velasco L, Perez-Vich B, Martinez JMF (2000) Inheritance of oleic acid content under controlled environment. In: Proceedings of the 15th international sunflower conference, Toulouse, 12–15 June 2000, pp 31–36

Velasco L, Dominguez J, Martinez JMF (2004) Registration of T589 and T2100 sunflower germplasms with modified tocopherol profiles. Crop Sci 44:361–362

Vera-Ruiz EM, Perez-Vich B, Martinez JMF, Velasco L (2005) Comparative genetic study of two sources of beta-tocopherol in sunflower. Helia 28(42):1–8

Vera-Ruiz EM, Velasco L, Leon AJ et al (2006) Genetic mapping of the *Tph1* gene controlling beta-tocopherol accumulation in sunflower seeds. Mol Breed 17(3):291–296

Warburton ML, Rauf S, Marek L, Hussain M, Ogunola O, Gonzalez JJS (2017) The use of crop wild relatives in maize and sunflower breeding. Crop Sci 57:1–14

Ye F, Liang Q, Li H, Zhao G (2015) Solvent effects on phenolic content composition, and antioxidant activity of extracts from florets of sunflower (*Helianthus annuus* L.). Ind Crop Prod 76:574–581

Zhu H, Li C, Gao C (2020) Applications of CRISPR–Cas in agriculture and plant biotechnology. Nat Rev Mol Cell Biol 21:661–677

Next-Generation Breeding for Nutritional Traits in Peanut

Priya Shah, Manish Pandey, Spurthi N. Nayak, Charles Chen, Sandip Bera, Chittaranjan Kole, and Naveen Puppala

Contents

Abstract

Peanut, an important oilseed legume, is largely grown in Asia, Africa, and the Americas. It is a source of vital nutrients and amino acids for optimal health. In addition to its health advantages, further enhancing the vitamin and protein content may aid in alleviating the issue of hidden hunger, particularly in Asian and African countries. Due to the availability of high-oleic peanut oil, consumers now have access to another inexpensive cooking oil that offers similar advantages to olive oil in terms of quality. Hence, there is a need to put immense focus on generating varieties of peanuts that are high in nutrients along with the yield, oil

P. Shah · M. Pandey
International Crops Research Institute for the Semiarid Tropics, Hyderabad, Telangana, India

S. N. Nayak
Department of Biotechnology, University of Agricultural Sciences, Dharwad, India

C. Chen
Department of Crop, Soil, & Environmental Sciences, Auburn University, Auburn, AL, USA

S. Bera
ICAR-Directorate of Groundnut Research, Junagadh, Gujarat, India

C. Kole
Institute of Nutraceutical Research, Clemson University, Clemson, SC, USA

N. Puppala (✉)
Agricultural Science Center at Clovis, New Mexico State University, Clovis – New Mexico, USA
e-mail: npuppala@nmsu.edu

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_15

content, and disease resistance. Development of nutrient-rich peanuts can be accelerated through marker-assisted selection and genomics-assisted breeding, followed by functional characterization of the candidate genes. In addition to current research on breeding, genetics, and genomics studies, this review also deliberates on the possible application of genetic engineering and genome editing to improve nutritional traits.

Keywords

Peanut · Functional compounds · Allergens · Marker-assisted selection · Genomic-assisted breeding · Genetic engineering · Genome editing

1 Introduction

Peanut (*Arachis hypogaea*, Family Fabaceae) is a valuable source of nutrition, most notably for those with a poor socio-economic conditions, as it is inexpensive. It is a rich source of amino acids comprised of protein, lipids like unsaturated fatty acids (UFA), and other components like dietary fibers and polysaccharides. According to a recent analysis, the approximate nutritional makeup of peanuts was 31–46% fat, 20.7–25.3% protein, and 21–37% carbohydrates. Peanut is also referred as "Poor man's almond" with its dietary properties not only as food but also for its functional components augmenting health benefits (Nayak et al. 2021).

Molecular markers have been used to improve peanut oil and oleic acid content (Sarvamangala et al. 2011; Shasidhar et al. 2020). Recent developments in genomics have paved the way to better understand the molecular mechanisms that are involved in improving the nutritional properties of peanut. In this regard, the reference genomes for diploid (Bertioli et al. 2016; Chen et al. 2016) and tetraploid (Bertioli et al. 2019; Chen et al. 2019; Zhuang et al. 2019) progenitors have made genome-wide single nucleotide polymorphism (SNP), variant genetic mapping, and genome-wide association studies (GWAS) in peanuts much easier. In a GWAS analysis of 120 genotypes from the US mini core collection, 24 quantitative trait loci (QTLs) for boron (B), two QTLs for copper (Cu), six QTLs for sodium (Na), three QTLs for sulfur (S), and one quantitative trait locus (QTL) for zinc (Zn) with 18.35–27.56% phenotypic variance expression (PVE) explained by each QTLs (PVE) were identified. Furthermore, mining genomic regions yielded 110 unrelated candidate genes. Interestingly, the key elemental/metal transporter gene *arahy.KQD4NT* (position5,413,913-5,417,353) was discovered on chromosome B04 (Zhang et al. 2019). Peanut germplasm has many genotypes in terms of oil content, which makes it possible to use GWAS to find genetic areas linked to health related (HR) traits in peanuts (Yol et al. 2017). In terms of oil content, several QTLs have been identified that regulate grain oil content, fatty acid composition, and protein content (Pandey et al. 2014b; Shasidhar et al. 2017; Wilson et al. 2017; Liu et al. 2020a). In the Tifrunner × GT-C20 recombinant inbred line (RIL) population, six QTLs were detected for oil content accounting for 3.0–10.2% PVE and nine QTLs for oil content accounting for 3.9–14.0% PVE percent using an RIL population (Sun Oleic

97R × NC94022) (Pandey et al. 2014b). On chromosome B03, a significant QTL (*qOCB3*) for oil content was discovered with 14.36% PVE and 3.9 logarithm of odds (LOD) (Huang et al. 2015). Later on three main QTLs for oil content were discovered in an advanced backcross population with 18.0–25.0% PVE (Wilson et al. 2017), and eight QTLs for oil content were identified in an RIL population (ICGV07368 × ICGV06420) with 5.67–22.11% PVE (Shasidhar et al. 2017). According to several studies, oil content has additive inheritance, which allows for the pyramiding of linked loci in the peanut breeding program (Fu et al. 2017; Shasidhar et al. 2017; Zhaoming et al. 2017). In an RIL population (TG 26 × GPBD 4), three QTLs explaining 1.5–10.2% PVE were detected (Sarvamangala et al. 2011).

Seven QTLs have been identified on five chromosomes (A04, A05, A08, B05, B06), accounting for 6.07–27.19% PVE (Xuhua 13 × Zhonghua 6), with one large and stable QTL (*qOCA08.1*) on A08 accounting for 10.14–27.19% PVE (Liu et al. 2020a). Similarly, 27 QTLs for oil content were discovered using a whole-genome resequencing technique, including a major and stable *qA05.1* QTL with a LOD range of 13.62–26.94 and 9.62–22.74% PVE. In an RIL population Zhonghua10×ICG12625, two major and stable QTLs *qOCB06* with 22.59% PVE and *qOCB10.1* with a PVE range 9.18–12.55% were observed across three environments (Guo et al. 2021).

Besides oil content, fatty acid components are other key quality factors linked to shelf-life of peanut products and has health benefits. A previous study revealed two mutations of the *fatty acid desaturase* (*FAD*) gene *FAD2A* and *FAD2B* from the A- and B-subgenomes, respectively, from high O/L (oleic/linoleic acid) genotypes. The *FAD* gene encodes enzymes that allow the transformation of oleic acid into linoleic acid in peanuts (Lopez et al. 2000; Chu et al. 2011). In addition, QTL analysis was performed using two RIL populations (SunOleic 97R × NC94022 and Tifrunner × GT-C20) to determine the proportional contribution of *FAD2* alleles to oil quality (Pandey et al. 2014b). This study found 21 QTLs with 1.04–42.33% PVE for oleic acid, linoleic acid, and oleic/linoleic acid ratio in the SunOleic 97R × NC94022 population and 23 (M-QTLs) with 3.63–28.98% PVE in the Tifrunner × GT-C20 population for oleic acid, linoleic acid, and O/L ratio. The influence of *FAD2* alleles on oleic and linoleic acid concentration was recently explored and identified QTLs for oleic acid (C18:1), linoleic acid (C18:2), and the ratio of oleic acid to linoleic acid (O/L) were positioned on linkage groups A03, A04, A09, B09, and B10. This study was based on phenotyping in seven environments, and it was found that Marker2575339 and Marker2379598 in B09 were associated with C18:1, C18:2, and O/L in all seven environments, whereas Marker4463600 and Marker4391589 in A09 were associated with C18:1, C18:2, and O/L in six environments. The results were verified in multiple genetic backgrounds using a high-density genetic map (Hu et al. 2018).

Four QTLs linked with oleic acid, linoleic acid, and the O/L ratio were discovered in the RIL population (TG 26 × GPBD 4), accounting for 1.4–9.7% PVE (Sarvamangala et al. 2011). In the RIL population (Zhonghua 10 × ICG12625), 10 QTLs (seven significant) accounting for 1.72–20.20% PVE were detected for six fatty acids (Huang et al. 2015). Another 20 significant QTLs with 10.3–78.6% PVE

with LOD values ranging from 3.7 to 191 have recently been discovered in the F_2 population (ICGV 06420 × SunOleic 95R) (Shasidhar et al. 2017). Another study reported a significant QTL on chromosome A09 in an RIL population TMV 2 × TMV 2-NLM that explains a 15.1% PVE for oleic acid (Hake et al. 2017). In a more recent study, four QTL clusters for saturated fatty acids were identified (palmitic, stearic, arachidic, behenic, and lignoceric acid) (Liu et al. 2019). On chromosome B04, 20 significant QTLs were detected on three QTL clusters (CLB04–1, CLB04–2, and CLB04–3), accounting for 10.77–41.89% PVE. Six QTLs for stearic, arachidic, and behenic acid were found in another QTL cluster (CLB06) on chromosome B06, with up to 20.32% PVE. Further investigation into these QTL clusters will aid in understanding fatty acid metabolism and discovering diagnostic markers that may be utilized in marker-assisted selection (MAS) to improve peanut cultivars.

Protein content is another significant characteristic that improves peanuts' nutritional quality as food and feed. Six QTLs for protein content have been found in an RIL population with 1.50–10.70% PVE and 2.87–3.63 LOD, while a recent study found one main QTL flanked by AhTE0003-AhTE0332 markers, associated with protein content on chromosome A10 with 26.4% PVE and 11.2 LOD (Sarvamangala et al. 2011; Hake et al. 2017). In a RIL population (Zhonghua 6 × Xuhua 13, 186 progenies), a recent study discovered nine additive QTLs for resveratrol content with 5.07–8.19% PVE and LOD 2.50–3.64 across four environments (Luo et al. 2021). Further research is needed using high-density genotyping and sequencing-based mapping, which will lead to fine mapping of genomic areas and candidate gene discovery, allowing for faster and more precise breeding of nutrition-rich peanut types. Apart from oil, the by-products of peanut contain many other functional compounds like proteins, fibers, polyphenols, antioxidants, vitamins, minerals, which can be added as a functional ingredient into many processed foods (Arya et al. 2016). Studies have indicated that peanuts are an excellent source of compounds like flavonoids, phenolic acids, resveratrol, and phytosterols in the kernels and shells (Nepote et al. 2002; Adhikari et al. 2019).

Peanuts have also been termed as functional food with numerous functional components like Coenzyme Q10, which protects the heart during the hypoxia conditions. Peanuts are also a good source of dietary fiber and provide a wide range of essential nutrients, including several B group vitamins, vitamin E, minerals such as iron, zinc, potassium and magnesium, antioxidant minerals (selenium, manganese and copper), plus other antioxidant compounds (such as flavonoids and resveratrol) (Gülçin 2010). There are attempts to profile the nutraceutical properties especially total antioxidant activity and total polyphenolic compounds in 60 groundnut genotypes varying in skin color, i.e., from pure white to purple (Nayak et al. 2020). The study indicated that peanut skin color has positive correlation with antioxidant activity as well as the total polyphenol content. Transposable marker based association analysis indicated 24 marker-trait association related to nutritional traits like fat, crude protein, ash, crude fiber, carbohydrate, total polyphenol content, and antioxidant activity and had QTL with PVE more than 10% (Nayak et al. 2020). In another study, the total flavonoid content was linked with four expressed sequence

tag – simple sequence repeat (EST-SSR) markers in a Chinese peanut germplasm of 57 accessions. The EST-SSRs were derived from ESTs coding for HSFs (GM2284), TiC20 (GM2156), agglutinin (GM2067) and mitochondrial outer membrane porin protein (GM1878) (Hou et al. 2017). Genetic variability for total phenolics, flavonoids, and antioxidant activity of testaless seeds in a RIL population VG 9514 × TAG 24 was used to identify QTLs (Mondal et al. 2015). The results indicated five QTLs for total flavonoid content, four QTLs for DPPH radical scavenging activity, and a single QTL for total phenolic content in six linkage groups. Of these 10 QTLs, six were clustered in the linkage group A02 and A03 (Mondal et al. 2015).

The major problem associated with peanut is the presence of antinutritional compounds and toxin contamination. The incidence of toxin contamination caused by fungus along the food chain significantly impacts the availability of healthy and safe food. From standing in the field through postharvest processing and storage, mycotoxin has been reported to harm one-fourth of the world's food crops (Wu 2007). This fungal toxin is mostly associated with three genera: *Aspergillus*, *Fusarium*, and *Penicillium* (Reddy et al. 2010). Aflatoxin produced by the genus *Aspergillus* can be found in various foods, including maize, peanuts, dried fruits, spices, milk, and its derivatives (Pandey et al. 2019c; Soni et al. 2020). Aflatoxin-contaminated food has significantly but negatively influenced human health (Atherstone et al. 2016). As a result, several countries have enacted stringent safety regulations and precise instructions to limit aflatoxin exposure to human health and welfare.

The European Union (EU) has imposed a rigorous guideline for aflatoxin-affected food consumption of 4 g/kg (European Commission-EC 2010). Similarly, the allowable limit for aflatoxin contamination in the United States is 20 g/kg (Wu 2007). Nonetheless, the EU's criteria are not practical to implement globally because many nations with lower GDPs and poor economies, particularly those in Africa and Asia, cannot comply with these regulations due to the higher expense of cultivation required to reach those standards. As a result, alternatives are required to keep aflatoxin levels below the safe limit, taking into account human health and wealth. Understanding the *Aspergillus* biology and the toxin it produces is one step toward improving crops with low aflatoxin contamination. New strategies to breed peanut cultivars with low aflatoxin contamination are required to ensure advantageous exports and healthy life.

Another problem with this crop is peanut allergy, a serious food allergy that affects 1–2% of the world's population and is the most common cause of anaphylaxis or death (Pandey et al. 2019b). It has been found that Australia, United States, Canada, Denmark, United Kingdom, and France are among the several other countries that are most affected by peanut allergens. Until recently, there was no accessible allergy vaccination or strategy for lowering allergenicity from peanut foods. There are 32 types of seed storage proteins in the peanut kernel, and 18 are allergenic proteins (Pele 2010).

Peanut allergens are divided into two categories: major and minor based on their lethal reactions detected with IgE, among this Ara h1, Ara h2, Ara h3, and Ara h6 are

classified as major allergens. These reactions can result in anaphylaxis or death. Around 617 kb from the cultivated peanut genome (cv. Florunner UF-439-16-1003-2) and 215 kb from a wild relative (*Arachis duranensis*; A genome) were sequenced in a previous study using the technique of bacterial artificial chromosome sequencing, and three Ara h 1, one Ara h 2, eight Ara h 3, and two Ara h 6 allergen coding genes were identified (Ratnaparkhe et al. 2014).

In addition, the A genome has 21 allergen-coding genes. Nine of these have already been detected in peanuts, with the remaining were homologous of the other crops (Chen et al. 2016). Recently, a monoclonal antibody-based sandwich ELISA approach has been effectively standardized on diverse peanut accessions, and significant allergens such as Ara h 1, Ara h 2, Ara h 3, Ara h 6, and Ara h 8 have been screened (Pandey et al. 2019b). Developing lower allergen-containing lines or cultivars is proclaimed to be a future breeding strategy (Pandey et al. 2019a).

Recently, studies conducted by the US FDA (Food and Drug Administration) revealed that peanut consumption in infants between the ages of 4 and 10 months, has reduced the fear of developing a severe peanut allergy, as it is stated that early peanut consumption is one of the paths to reducing severe peanut allergy (https://www.fda.gov/food/cfsan-constituentupdates/fda-acknowledges-qualified-health-claim-linking-ear). According to a health claim regarding allergic reactions to peanuts, the FDA has established the connection between the use of food sources containing powdered peanuts and a reduced risk of developing an allergy, with the logical proof that is acceptable and has advised realizing offices to provide accurate data for nourishments to avoid deceiving consumers. Furthermore, the FDA will examine peanut-containing foods and issue a qualified health certificate to reduce the concern of peanut allergy (https://www.fda.gov/media/107357/download). Peanut lines with low allergen concentrations can generate vaccines or therapeutic medicines, reducing the risk of peanut allergies.

2 Molecular Breeding and Biotechnology for Nutritional Traits

The genome-wide association studies (GWAS) allow us to unravel the trait of interest in a diverse panel with high precision; therefore, this approach can be useful for studying the genetics of nutritional traits in peanuts. This strategy can be effective for exploring the genetics of HR characteristics in peanuts because association studies allow us to unravel the trait of interest in a broad panel with high precision. A panel of 292 lines was used to identify 12 linked oil content markers, including one highly stable connection (AGGS1014_2) with 9.94% PVE (Liu et al. 2020b). Similarly, GWAS was used to investigate the genetic basis of HR traits in a reference set of 300 diverse global peanut collections, resulting in the discovery of 24 marker-trait associations (MTAs), including two MTAs for oleic acid content (16.42–20.8% PVE), 22 MTAs for O/L ratio (13.7–47.45% PVE), 25 MTAs for oil content (5.84% (gnPt-714,399) to 40.37% (TC4G10) PVE,

11 MTAs for protein content with 11.63–36.1% PVE, and 1 MTA for zinc content with 15.63% PVE (Pandey et al. 2014a).

Molecular breeding has been successfully utilized to improve crops' nutritional content and improve nutritional traits. Peanut oil has a higher unsaturated fatty acid (UFA) to saturated fatty acid (SFA) ratio than palm and coconut oil, making it a healthier cooking oil. Peanut oil's high oleic acid (HOA) content aids in the reduction of low-density lipoprotein cholesterol (LDLC) levels and the risk of cardiovascular disease (CVD). To improve peanut lines with HOA, a healthier choice for eradicating malnutrition in rural and tribal communities, breeding for improved peanut lines with HOA is required.

Florida Agricultural Experiment Station in the United States released SunOleic 97R, a multiline cultivar with 81.8% oleic acid content (Gorbet and Knauft 2000). It is made up of three breeding lines descended from a cross between F435–2-2-E-2-l-b4-E-b2-b3-l-E (high oleic) and "Sunrunner" (F519–9), with the latter serving as the recurrent parent (Norden et al. 1985; Gorbet and Knauft 2000). With the discovery of diagnostic markers for HOA, peanut breeding techniques, including marker-assisted selection (MAS) and marker-assisted backcrossing (MABC) have emerged as viable options for crop improvement. As a result, two simultaneous backcross procedures were employed to increase oleic acid levels in Tifguard (nematode-resistant cultivar), utilizing linked markers for HOA and nematode resistance, with Florida-07 and Georgia-02C as HOA donors (Chu et al. 2011). Phenotyping was done on progenies of BC_3F_2 that demonstrated homozygosity for HOA and nematode resistance alleles to supplement the precision of MAS. SunOleic 95R (HOA line) mutant alleles of *FAD2A* and *FAD2B* were transferred into ICGV 06420, ICGV 06142, and ICGV 06110 using MABC and MAS (Janila et al. 2016).

Consequently, 27 lines with 53–58% oil content and 80% oleic acid were improved, while 28 lines with 42–50% oil content and 80% oleic acid were enhanced. Backcross lines with HOA were created, and multilocation yield studies were conducted. As a result, the varieties "Girnar 4" and "Girnar 5" were identified and launched in India in 2020 as the best-performing varieties with HOA content (Nawade et al. 2016; Bera et al. 2019). To increase the oleic acid content of four Chinese peanut cultivars (Yuanza 9102, Yuhua 9326, Yuhua 9327, and Yuhua 15), competitive allele-specific PCR (KASP) assay-based MABC was used to detect *FAD2A* mutations (Huang et al. 2019). Consequently, 24 HOA lines with similar agro-morphological traits to recurrent parents (with 79.49–92.31% genome recovery) have been produced and are undergoing multilocation experiments in preparation for possible release. The MABC technique was used to deploy GJGHPS 1, GJG 9, and 20 to boost HOA content and foliar disease resistance (FDR) (Shasidhar et al. 2020).

As a result, in the BC_3F_7 generation, >50 FDR ILs (introgression lines) and > 80 high oleic ILs were produced and carried forward for seed multiplication. For HOA and FDR, Kadiri 6, Dh86, ICGV 00351, and ICGV 87846 were used to generate >200 ILs (BC_3F_4) (ICRISAT unpublished). More than 200 pyramided lines were created using the above ILs by incorporating HOA and FDR into all six kinds and are being tested, evaluated, and released. In addition, more than 300 HOA breeding lines (F_3-F_7 and BC_3F_3-BC_3F_7) have been produced in the background of high-yielding varieties such

as GG22, GG20, GJG32, Kadiri-6, DH86, DH256, DH257, KadiriLepakshi, TG37A, TKG19A, TG51, TG81, JL 501, Girnar 2, NRCGCS268, and NRCGCS257. They are in different stages of testing (ICAR-Directorate of Groundnut Research, India).

Mutation breeding is also used in improving peanut cultivars to develop cultivars with high oil content. The genotypes Minhua 8 and Minhua 6 were exposed to gamma-ray and EMS-based mutagenesis, respectively, to develop new high oleate lines in peanuts (Zhuang et al. 2019). Until then all the HOA lines were developed by the natural mutation at *ahFAD2A* and *ahFAD2B*loci. As a result, three Minhua 8 varieties and four Minhua 6 cultivars were developed with HOA and improved agronomic performance. Two of them are reportedly undergoing multilocation trials in preparation for further testing. MABC was recently used to increase the oleic acid content of two elite genotypes, GPBD 4 and G-252, which have high productivity, oil content, and resistance to late leaf spot (LLS) and rust diseases. Only the mutant allele at *AhFAD2B* was transferred from the donor SunOleic 95R because both the recurrent parents already had the mutant allele at *AhFAD2A* (80.6%). In $BCnF_2$ generations, three rounds of backcrossing with foreground selection using allele-specific PCR and the KASP assay revealed a substantial number of plants homozygous for the mutant allele at *AhFAD2B*.

Targeted mutations can be created to address human needs shortly, thanks to the advances in genome-editing technology. The peanut crop has a significant potential to provide customers with extremely nutritious items that will not only assist in addressing the issue of malnutrition but will also help to provide high energy and nutrition while consuming fewer food products. For breeding more nutritious peanut cultivars, next-generation genetic improvement technologies such as genomic selection and genome editing should be investigated. Most importantly, gaining a better understanding of nutritional traits through precision phenotyping and sequencing will aid in identifying the causal genes that make this crop so nutrient-dense. The promotion and acceptance of nutrition-dense peanut cultivars should also be prioritized to guarantee that the benefits reach the farmers, industry, and consumers.

3 Genetic Transformation and Gene Editing in Peanut

Increasing food production by cultivating high-yielding, high-quality crops is critical to ensure food security. Various genetic transformation and gene-editing tools have recently been developed. These strategies have been used in various scientific sectors, and their development has improved nutritional quality (Yuan et al. 2019; Biswas et al. 2022). Plant breeders use diverse gene transformation strategies to impart biologically essential features and produce high-yielding crops with good nutritional quality and other traits for various reasons. Two genetic transformation approaches, *Agrobacterium*-mediated and biolistic strategies, are essential to develop commercially relevant traits and alter plant genetic resources. In a particle bombardment approach, immature peanut cotyledons were used as explants (Deng et al. 2001). Exogenous DNA segments are freed in the nucleus and can be integrated into chromosomal DNA by illegitimate or homologous recombination

processes, which depend solely on cellular components. Micro-projectile bombardment of peanut tissue has been reported in several publications, including bombarding leaflets from mature embryos (Clemente et al. 1992) and the somatic embryogenesis regeneration system (Deng et al. 2001).

Peanut is thought to be difficult to culture and genetically modify. Many species either resist or have limited effectiveness with this gene transformation technique (Estrada-Navarrete et al. 2006). Successful genetic transformation by *Agrobacterium*-mediated technique has been described in peanut research employing various explant sources such as de-embryonated cotyledons (Sharma and Anjaiah 2000), embryo axes (Grabiele et al. 2012), and cotyledons (Bhatnagar et al. 2010). Cotyledonary nodes (CNs) are more able to regenerate than other explant sources. *Agrobacterium*-mediated transformation in peanuts has been reported using a variety of explant sources, including a cotyledonary node, de-embryonated cotyledon, leaflet, immature leaves (Mehta et al. 2013; Marka and Nanna 2018).

Genetic transformation is useful for studying plant gene function, although restricted transformation capability limits its influence in some systems. *Agrobacterium tumefaciens*-mediated transformation has been used to research gene function in legumes. Furthermore, increased availability and cost-effectiveness are two benefits that can be obtained using this gene transfer technology. Genome-editing technologies are very new and widely used in plant research to modify genes linked to nutritional quality, yield, disease resistance, and drought resilience (Kaniganti et al. 2022).

Zinc-finger nucleases (ZFNS), transcription activator-like effector nucleases (TALEN), and clustered regularly interspaced short palindromic repeats system (CRISPR/Cas9), for example, are currently used widely by researchers and are helping to boost productivity and production. CRISPR/Cas9, a gene-editing tool that is simple and inexpensive, is frequently used in plant breeding programs. Peanuts are an important oil-producing allotetraploid crop. As a result, peanut gene editing technology could boost the amount of oleic acid in edible peanut oil.

Genome-editing methods based on site-directed endonucleases capable of causing chromosomal double-strand breaks (Kim et al. 1996; Christian et al. 2010; Jinek et al. 2012) can assist overcome the limits of traditional breeding and speed up the production of better crops. Double-stranded breaks (DSBs) can deliver targeted disease resistance and genome alterations to improve agronomic parameters such as yield and nutritional content by harnessing the natural cellular DNA repair process. The most promising approach for developing an improved genotype has been genome editing technology. Peanut genome editing techniques targeted HR characteristics such as high oleic acid content and allergens(Yuan et al. 2019; Biswas et al. 2022).

It is anticipated that using the revolutionary biotechnological tool as genome editing will enhance crops' nutritive and practical quality. Rapidly developing genome editing technology might make it possible to create grain legumes with increased protein levels without inserting foreign genes. The development of crop plants with enhanced and novel traits of economic and nutritional importance has been made possible by genome editing technologies. Various crop plants have

improved quality traits, such as increased fragrance and low gluten, starch, or oleic acid contents (Ku and Ha 2020). Genetic engineering is applied to improve the nutritional quality of plants, and increase stability during processing and storage, among others (Tien Lea et al. 2016). In this context, a significant amount of work is needed to pinpoint the genes involved in fatty acid composition metabolism and create a system of gene-specific markers to enhance crop nutrition. The use of molecular markers can greatly aid breeding operations by revealing the precise position of the genomic region/quantitative trait locus (QTL) that controls the nutrient content trait. The markers that are associated with QTL have the potential to be used across breeding material for identification and introgression (Gaikwad et al. 2020).

Plant breeders have identified crops with commercially important features through germplasm screening. Beans, particularly peanuts, were recognized as having a high oleic acid concentration. F435, a high oleate spontaneous mutant line, contains 80% oleic acid and 2% linoleic acids (Norden et al. 1987). Two types of mutations were found in this mutant line: "G" to "A" substitution occur at 448 base pair (bp) in the *ahFAD2A* gene that result in a missense mutation whereas there occur an "A" insertion between 441 and 442 bp (441442insA) in the *ahFAD2B* gene (Jung et al. 2000). Using this mutant line (F435) as one of the parents in normal peanut breeding helps to improve essential traits and since then, many cultivars with high oleic acid to linoleic acid ratio (O/L) have been created (Chu et al. 2009). Increasing the amount of oleic acid in the peanut genome improves shelf-life and provides health benefits.

Peanut oil includes roughly 12 fatty acids, with oleic acid, a monounsaturated fatty acid (36–67%), and linoleic acid, a polyunsaturated fatty acid, accounting for nearly 80% of the total (15–43%). Furthermore, palmitic acid, a saturated fatty acid, accounts for roughly 10% of the total, with the remaining 10% made up of up to nine additional fatty acids; (Janila et al. 2016). The presence of various fatty acids, such as saturated fatty acids, monounsaturated fatty acids, and polyunsaturated fatty acids, in peanut oil, determines the nutritional quality, flavor, and shelf-life of peanut seeds and products. Due to consumer and commercial benefits such as antioxidation and long shelf-life, increasing the concentration of oleic acid in peanut seeds is one of the key goals in peanut breeding.

The fatty acid desaturases encoded by the homologous *ahFAD2A* and *ahFAD2B* genes are the primary enzymes for converting oleic acid to linoleic acid, easily oxidized. So far, all high oleic acid peanut variants have resulted from spontaneous mutations in both genes. Inducing mutations in the DNA of other superior cultivars could help speed up the introduction of this desirable characteristic. Using peanut protoplasts and hairy root cultures as models, a gene-editing strategy based on CRISPR/Cas9 technology has been used to create de novo mutations in the *ahFAD2* genes.

There have been reports of a novel G451T mutation in the coding area of the *Arachis hypogaea FAD2B* gene in the peanut caused by CRISPR/Cas9-based gene editing (Yuan et al. 2019). Using freshly established gene-editing technology, peanut breeders will increase the oleic acid to linoleic acid ratio (O/L) ratio. Several

high-oleate (HO) peanut cultivars have been developed in China, and nations such as Argentina, Australia, Brazil, Israel, Japan, and South Africa are also developing high-oleate peanut products for human consumption (Janila et al. 2016). The target region was a hotspot of a natural mutation in these genes. Suitable sgRNAs were created and cloned into a CRISPR/Cas9 expression plasmid.

Three mutations were discovered due to CRISPR/Cas9 activity: G448A in *ahFAD2A*, 441442insA, and G451T in *ahFAD2B*. The G448A and 441442insA mutations are in current high oleate types, but the G451T mutation is new. Because natural mutations in the *ahFAD2A* gene are more common than in the *ahFAD2B* gene in *A. hypogaea* var. *hypogaea*, the mutations induced in *ahFAD2B* by gene editing may be useful in developing high oleate lines with a variety of genetic backgrounds after validation of oleic acid content in transformed lines. The G448A mutation in *ahFAD2A* is another advantage of high oleic acid oil content. The CRISPR-based genome editing generated mutations in peanuts, and this CRISPR/Cas9 technology could be valuable in peanut breeding efforts. The desaturase enzyme encodes the *ahFAD2A* and *ahFAD2B* genes (Jung et al. 2000). This enzyme is extremely important in converting oleic acid to linoleic acid. Hence, gene editing to lower linoleic acid and boost oleic acid in the peanut genome significantly impacts future breeding programs (Yuan et al. 2019).

Similarly, the protoplast culture has been used to disrupt the *Ara h2* gene, a major allergen gene, with the help of an endogenous tRNA-processing system. Here, multiplex CRISPR/Cas9 genome editing was performed in peanut protoplasts. Finally, several edits were identified in the target gene after polyethylene glycol (PEG)-mediated transformation in protoplasts with a Cas9 and sgRNA-containing vector through deep-sequencing analysis. This strategy provides an efficient pipeline to develop gene-editing constructs for various genes or peanut transformation. These findings demonstrated that PEG-mediated protoplast transformation system could serve as a rapid and effective tool for transient expression assays and sgRNA validation in peanuts. Once optimized, stable transformants can be developed using *Agrobacterium*-mediated transformation or alternative delivery systems. However, an efficient gene-editing platform in peanuts needs to be established to assist in basic research in understanding gene functions and molecular pathways and to help speed up breeding programs in developing peanuts with improved yield, quality, and tolerance to various abiotic and biotic stresses. Indeed, further optimization of the CRISPR/Cas9 system in peanuts can be explored using other editing techniques, including allele replacement, to widen the target traits and speed up the breeding progress (Biswas et al. 2022).

4 Conclusions

Peanuts are cultivated in the continents of Asia and Africa, primarily in the semi-arid zones where malnutrition is worrisome. Numerous high-oleic peanut genotypes have been developed worldwide using conventional breeding techniques, but little attention has been given to other nutritional components. The peanut crop has enormous

potential to provide consumers with highly nutrient-rich products that will not only help solve the problem of malnutrition but also provide high energy and nutrition by consuming fewer food products. Next-generation genetic improvement techniques, including genomic selection and genome editing, will prove immensely promising for developing more nutrient-dense peanut cultivars. These techniques are highly applicable for functional validation and are widely used in plant research to modify genes linked to nutritional quality, yield, disease resistance, and drought resilience. The most critical step will be identifying the causal genes that make this crop so nutrient-rich by developing a greater understanding of nutritional features through precision phenotyping and sequencing. Additionally, it is important to prioritize adopting and promoting nutrient-dense peanut cultivars to guarantee that farmers, businesses, and consumers reap the rewards.

References

Adhikari B, Dhungana SK, Ali MW, Adhikari A, Kim I-D, Shin D-H (2019) Antioxidant activities, polyphenol, flavonoid, and amino acid contents in peanut shell. J Saudi Soc Agric Sci 18(4): 437–442

Arya SS, Salve AR, Chauhan S (2016) Peanuts as functional food: a review. J Food Sci Technol 53: 31–41

Atherstone C, Grace D, Lindahl JF, Kang'ethe EK, Nelson F et al (2016) Assessing the impact of aflatoxin consumption on animal health and productivity. Afr J Food Agric Nutr Dev 16(3): 10949–10966

Bera SK, Kamdar JH, Kasundra SV, Patel SV, Jasani MD, Maurya AK et al (2019) Steady expression of high oleic acid in peanut bred by marker-assisted backcrossing for fatty acid desaturase mutant alleles and its effect on seed germination along with other seedling traits. PLoS One 14(12):e0226252

Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EKS et al (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet 48(4):438–446

Bertioli DJ, Jenkins J, Clevenger J, Dudchenko O, Gao D, Seijo G et al (2019) The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. Nat Genet 51(5):877–884

Bhatnagar M, Prasad K, Bhatnagar-Mathur P, Lakshmi Narasu M, Waliyar F, Sharma KK (2010) An efficient method for the production of marker-free transgenic plants of peanut (*Arachis hypogaea* L.). Plant Cell Rep 29(5):495–502

Biswas S, Wahl NJ, Thomson MJ, Cason JM, McCutchen BF, Septiningsih EM (2022) Optimization of protoplast isolation and transformation for a pilot study of genome editing in peanut by targeting the allergen gene Ara h 2. Int J Mol Sci 23(2):837

Chen X, Li H, Pandey MK, Yang Q, Wang X, Garg V et al (2016) Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens. Proc Natl Acad Sci U S A 113(24):6785–6790

Chen X, Lu Q, Liu H, Zhang J, Hong Y, Lan H et al (2019) Sequencing of cultivated peanut, *Arachis hypogaea*, yields insights into genome evolution and oil improvement. Mol Plant 12(7): 920–934

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A et al (2010) Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186(2):757–761

Chu Y, Holbrook CC, Ozias-Akins P (2009) Two alleles of ahFAD2B control the high oleic acid trait in cultivated peanut. Crop Sci 49(6):2029–2036

Chu Y, Wu CL, Holbrook CC et al (2011) Marker-assisted selection to pyramid nematode resistance and the high oleic trait in peanut. Plant Genome 4:110–117. https://doi.org/10.3835/plantgenome2011.01.0001

Clemente TE, Robertson D, Isleib TG, Beute MK, Weissinger AK (1992) Evaluation of peanut (*Arachis hypogaea* L.) leaflets from mature zygotic embryos as recipient tissue for biolostic gene transfer. Transgenic Res 1(6):275–284

Deng XY, Wei ZM, An HL (2001) Transgenic peanut plants obtained by particle bombardment via somatic embryogenesis regeneration system. Cell Res 11(2):156–160

Estrada-Navarrete G, Alvarado-Affantranger X, Olivares J-E, Díaz-Camino C, Santana O, Murillo E et al (2006) *Agrobacterium rhizogenes* transformation of the Phaseolus spp.: a tool for functional genomics. Mol Plant-Microbe Interact 19(12):1385–1393

European Commission-EC (2010) Commission regulation (EU) no 165/2010 of 26 February 2010, amending regulation (EC) no 1881/2006 setting maximum levels for certain contaminants in foodstuffs as regards aflatoxin. OJEU 50:8–12

Fu Y, Zhang D, Gleeson M, Zhang Y, Lin B, Hua S et al (2017) Analysis of QTL for seed oil content in *Brassica napus* by association mapping and QTL mapping. Euphytica 213(1):1–15

Gaikwad KB, Rani S, Kumar M, Gupta V, Babu PH, Bainsla NK et al (2020) Enhancing the nutritional quality of major food crops through conventional and genomics-assisted breeding. Front Nutr 7:533453

Gorbet DW, Knauft DA (2000) Registration of Sun oleic 97R'Peanut. Crop Sci 40(4):1190

Grabiele M, Chalup L, Robledo G, Seijo G (2012) Genetic and geographic origin of domesticated peanut as evidenced by 5S rDNA and chloroplast DNA sequences. Plant Syst Evol 298(6):1151–1165

Gülçin İ (2010) Antioxidant properties of resveratrol: a structure–activity insight. Innov Food Sci Emerg Technol 11(1):210–218

Guo J, Liu N, Li W, Wu B, Chen H, Huang L et al (2021) Identification of two major loci and linked marker for oil content in peanut (*Arachis hypogaea* L.). Euphytica 217(2):1–11

Hake AA, Shirasawa K, Yadawad A, Sukruth M, Patil M, Nayak SN et al (2017) Mapping of important taxonomic and productivity traits using genic and non-genic transposable element markers in peanut (*Arachis hypogaea* L.). PLoS One 12(10):e0186113

Hou M, Mu G, Zhang Y, Cui S, Yang X, Liu L (2017) Evaluation of total flavonoid content and analysis of related EST-SSR in Chinese peanut germplasm. Crop Breed Appl Biotechnol 17:221–227

Hu XH, Zhang SZ, Miao HR, Cui FG, Shen Y, Yang WQ et al (2018) High-density genetic map construction and identification of QTLs controlling oleic and linoleic acid in peanut using SLAF-seq and SSRs. Sci Rep 8(1):1–10

Huang L, He H, Chen W, Ren X, Chen Y, Zhou X et al (2015) Quantitative trait locus analysis of agronomic and quality-related traits in cultivated peanut (*Arachis hypogaea* L.). Theor Appl Genet 128(6):1103–1115

Huang B, Qi F, Sun Z, Miao L, Zhang Z, Liu H, et al. et al (2019) Marker-assisted backcrossing to improve seed oleic acid content in four elite and popular peanut (*Arachis hypogaea* L.) cultivars with high oil content. Breed Sci 69:18107

Janila P, Pandey MK, Shasidhar Y, Variath MT, Sriswathi M, Khera P et al (2016) Molecular breeding for introgression of fatty acid desaturase mutant alleles (*ahFAD2A* and *ahFAD2B*) enhances oil quality in high and low oil containing peanut genotypes. Plant Sci 242:203–213

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science 337(6096):816–821

Jung S, Swift D, Sengoku E, Patel M, Teule F, Powell G et al (2000) The high oleate trait in the cultivated peanut [*Arachis hypogaea* L.]. I. Isolation and characterization of two genes encoding microsomal oleoyl-PC desaturases. Mol Gen Genet 263(5):796–805

Kaniganti S, Chaitanya AK, Shah P, Magar ND, Jamedar HVR, Sajjad M. (2022) Status and prospects of genetic improvement of grain legumes using in vitro techniques: a review. Available at SSRN 4100545

Kim Y-G, Cha J, Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci 93(3):1156–1160

Ku H-K, Ha S-H (2020) Improving nutritional and functional quality by genome editing of crops: status and perspectives. Front Plant Sci 11:577313

Liu N, Chen H, Huai D, Xia F, Huang L, Chen W et al (2019) Four QTL clusters containing major and stable QTLs for saturated fatty acid contents in a dense genetic map of cultivated peanut (*Arachis hypogaea* L.). Mol Breed 39(2):1–14

Liu N, Guo J, Zhou X, Wu B, Huang L, Luo H et al (2020a) High-resolution mapping of a major and consensus quantitative trait locus for oil content to a~ 0.8-Mb region on chromosome A08 in peanut (*Arachis hypogaea* L.). Theor Appl Genet 133(1):37–49

Liu N, Huang L, Chen W, Wu B, Pandey MK, Luo H et al (2020b) Dissection of the genetic basis of oil content in Chinese peanut cultivars through association mapping. BMC Genet 21(1):1–12

Lopez Y, Nadaf HL, Smith OD, Connell JP, Reddy AS, Fritz AK (2000) Isolation and characterization of the Δ12-fatty acid desaturase in peanut (*Arachis hypogaea* L.) and search for polymorphisms for the high oleate trait in Spanish market-type lines. Theor Appl Genet 101(7):1131–1138

Luo H, Guo J, Yu B, Chen W, Zhang H, Zhou X et al (2021) Construction of ddRADseq-based high-density genetic map and identification of quantitative trait loci for trans-resveratrol content in Peanut seeds. Front Plant Sci 12:644402

Marka R, Nanna RS (2018) Optimization of factors affecting agrobacterium-mediated genetic transformation in groundnut (*Arachis hypogaea* L.). Adv Plants Agric Res 8(3):275

Mehta R, Radhakrishnan T, Kumar A, Yadav R, Dobaria JR, Thirumalaisamy PP et al (2013) Coat protein-mediated transgenic resistance of peanut (*Arachis hypogaea* L.) to peanut stem necrosis disease through *agrobacterium*-mediated genetic transformation. Indian J Virol 24(2):205–213

Mondal S, Phadke RR, Badigannavar AM (2015) Genetic variability for total phenolics, flavonoids and antioxidant activity of testaless seeds of a peanut recombinant inbred line population and identification of their controlling QTLs. Euphytica 204:311–321

Nawade B, Bosamia TC, Thankappan R, Rathnakumar AL, Kumar A, Dobaria JR et al (2016) Insights into the Indian peanut genotypes for ahFAD2 gene polymorphism regulating its oleic and linoleic acid fluxes. Front Plant Sci 7:1271

Nayak SN, Hebbal V, Bharati P et al (2020) Profiling of nutraceuticals and proximates in peanut genotypes differing for seed coat color and seed size. Front Nutr 7:45

Nayak SN, Aravind B, Malavalli SS, Sukanth BS, Poornima R, Bharati P et al (2021) Omics technologies to enhance plant based functional foods: An overview. Front Genet 12:742095

Nepote V, Grosso N, Guzman CA (2002) Extraction of antioxidant components from peanut skins. Grasas Aceites 53:391–395

Norden AJ, Gorbet DW, Knauft DA (1985) Registration of 'Sunrunner' peanut. Crop Sci 25(6):1126

Norden AJ, Gorbet DW, Knauft DA, Young CT (1987) Variability in oil quality among peanut genotypes in the Florida breeding program. Peanut Sci 14(1):7–11

Pandey MK, Upadhyaya HD, Rathore A, Vadez V, Sheshshayee MS, Sriswathi M et al (2014a) Genomewide association studies for 50 agronomic traits in peanut using the 'reference set'comprising 300 genotypes from 48 countries of the semi-arid tropics of the world. PLoS One 9(8):e105228

Pandey MK, Wang ML, Qiao L, Feng S, Khera P, Wang H et al (2014b) Identification of QTLs associated with oil content and mapping FAD2 genes and their relative contribution to oil quality in peanut (*Arachis hypogaea* L.). BMC Genet 15(1):1–14

Pandey AK, Sudini HK, Upadhyaya HD, Varshney RK, Pandey MK (2019a) Hypoallergen peanut lines identified through large-scale phenotyping of global diversity panel: providing hope toward addressing one of the major global food safety concerns. Front Genet 10:1177

Pandey AK, Varshney RK, Sudini HK, Pandey MK (2019b) An improved enzyme-linked immunosorbent assay (ELISA) based protocol using seeds for detection of five major peanut allergens Ara h 1, Ara h 2, Ara h 3, Ara h 6, and Ara h 8. Front Nutr 6:68

Pandey MK, Kumar R, Pandey AK, Soni P, Gangurde SS, Sudini HK et al (2019c) Mitigating aflatoxin contamination in groundnut through a combination of genetic resistance and post-harvest management practices. Toxins 11(6):315

Pele M (2010) Peanut allergens. Rom Biotechnol Lett 15(2):5204–5212

Ratnaparkhe MB, Lee T-H, Tan X, Wang X, Li J, Kim C et al (2014) Comparative and evolutionary analysis of major peanut allergen gene families. Genome Biol Evol 6(9):2468–2488

Reddy KRN, Salleh B, Saad B, Abbas HK, Abel CA, Shier WT (2010) Erratum: An overview of mycotoxin contamination in foods and its implications for human health (toxin reviews (2010) 29 (326)). Toxin Rev 29(3–4)

Sarvamangala C, Gowda MVC, Varshney RK (2011) Identification of quantitative trait loci for protein content, oil content and oil quality for groundnut (*Arachis hypogaea* L.). Field crops res 122(1):49–59

Sharma KK, Anjaiah V (2000) An efficient method for the production of transgenic plants of peanut (*Arachis hypogaea* L.) through *agrobacterium tumefaciens*-mediated genetic transformation. Plant Sci 159(1):7–19

Shasidhar Y, Vishwakarma MK, Pandey MK, Janila P, Variath MT, Manohar SS et al (2017) 'Molecular mapping of oil content and fatty acids using dense genetic maps in groundnut (*Arachis hypogaea* L.). Front Plant Sci 8:794

Shasidhar Y, Variath MT, Vishwakarma MK, Manohar SS, Gangurde SS, Sriswathi M et al (2020) Improvement of three popular Indian groundnut varieties for foliar disease resistance and high oleic acid using SSR markers and SNP array in marker-assisted backcrossing. Crop J 8(1):1–15

Soni P, Gangurde SS, Ortega-Beltran A, Kumar R, Parmar S, Sudini HK et al (2020) Functional biology and molecular mechanisms of host-pathogen interactions for aflatoxin contamination in groundnut (*Arachis hypogaea* L.) and maize (*Zea mays* L.). Front Microbiol 11:227

Tien Lea D, Duc Chua H, Quynh Lea N (2016) Improving nutritional quality of plant proteins through genetic engineering. Curr Genomics 17(3):220–229

Wilson JN, Chopra R, Baring MR, Selvaraj MG, Simpson CE, Chagoya J et al (2017) Advanced backcross quantitative trait loci (QTL) analysis of oil concentration and oil quality traits in peanut (*Arachis hypogaea* L.). Trop Plant Biol 10(1):1–17

Wu F (2007) Measuring the economic impacts of fusarium toxins in animal feeds. Anim Feed Sci Technol 137(3–4):363–374

Yol E, Üstün R, Gölükçü M, Uzun B (2017) Oil content, oil yield and fatty acid profile of groundnut germplasm in mediterranean climates. Journal of the American Oil Chemists' Society 94:787–804. https://doi.org/10.1007/s11746-017-2981-3

Yuan M, Zhu J, Gong L, He L, Lee C, Han S et al (2019) Mutagenesis of FAD2 genes in peanut with CRISPR/Cas9 based gene editing. BMC Biotechnol 19(1):1–7

Zhang H, Wang ML, Schaefer R, Dang P, Jiang T, Chen C (2019) GWAS and coexpression network reveal Ionomic variation in cultivated Peanut. J Agric Food Chem 67(43):12026–12036

Zhaoming Q, Xiaoying Z, Huidong Q, Dawei X, Xue H, Hongwei J et al (2017) Identification and validation of major QTLs and epistatic interactions for seed oil content in soybeans under multiple environments based on a high-density map. Euphytica 213(8):1–14

Zhuang W, Chen H, Yang M, Wang J, Pandey MK, Zhang C et al (2019) The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nat Genet 51(5):865–876

Genomic Designing for Nutraceuticals in *Brassica juncea*: Advances and Future Prospects

Aakanksha, Bal Govind Yadav, Shikha Mathur, Satish Kumar Yadava, and Nirala Ramchiary

Contents

Aakanksha · S. Mathur
Department of Genetics, University of Delhi South Campus, New Delhi, India

B. G. Yadav
International Centre for Genetic Engineering and Biotechnology, New Delhi, India

S. K. Yadava (✉)
Centre for Genetic Manipulation of Crop Plants, University of Delhi South Campus, New Delhi, India
e-mail: yadav24@gmail.com

N. Ramchiary (✉)
School of Life Sciences, Jawaharlal Nehru University, New Delhi, India
e-mail: nramchiary@jnu.ac.in

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_16

Abstract

Mustard (*Brassica juncea*) is an important crop of the Indian subcontinent and has been used as oilseed and as a condiment for a long time. Its importance as a nutraceutical crop is growing rapidly as it is a rich source of polyunsaturated fatty acids that have long been associated with cardioprotective activities, while mustard extract has been reported to harbor anticancerous properties. Mustard leaves and oil cake that is left after oil extraction are also excellent sources of antioxidants, minerals, and vitamins. However, mustard oil also contains glucosinolates and erucic acid in large amounts that are considered antinutritional as they have been implicated in goiter and cardiac lesions. The major objectives for the mustard breeding programs across the world therefore focus on enhancing the nutritional and functional properties with a concomitant overall increase in the seed oil percentage. As the nutraceutical properties of mustard are being recognized and are supported by many research studies based on QTL mapping, GWAS, genomic selection, etc., the breeding programs now also include the traits with potential nutraceutical applications. However, limited germplasm lines

with high therapeutic values, genetic bottleneck due to continuous selection for oil and yield, along with a low preference for varieties with possible nutraceutical applications by farmers, are some of the major hurdles in the development of mustard varieties with high nutritional and pharmaceutical values. Modern technologies like high-throughput genotyping, sequencing, and genome editing through the novel CRISPAR/CAS9 system hold a great promise in overcoming these impediments to achieving the targets of genetic improvement of nutraceutical properties of *B. juncea*.

Keywords

Brassica juncea · Nutraceutical · Sinigrin · Oil quality · Erucic acid · Biofortification · Biopharma · Omics · Molecular markers

1 Introduction

1.1 Agricultural Importance of Indian Mustard

Brassica juncea (AABB, 2n = 36) or Indian mustard is an oilseed crop majorly cultivated in South Asia, and India ranks fourth in its production worldwide. It is an amphidiploid species derived from the natural hybridization of diploid progenitor species *B. rapa* (AA, n = 10) and *B. nigra* (BB, n = 8). Based on cytological and molecular evidence, it is considered to have originated about 4000 years ago. It is an annual herb, adapted to diverse climatic conditions, and grows on varying soil types. Mustard is a crop of temperate region that prefers moderate temperature and is largely grown in the rabi season. The crop is mainly self-pollinated; however, cross-pollination to the extent of about 15–30% has also been reported (Rakow and Woods 1987).

Considerable variation exists in *B. juncea* for different important agro-morphological traits (Singh et al. 1980; Khulbe et al. 2000; Pant and Singh 2001) in collections from different agro-climatic zones. The gene pools of *B. juncea* can be divided into the East European and the Indian gene pools (Srivastava et al. 2001; Pradhan and Pental 2011). The germplasm of *B. juncea* has been utilized extensively in different breeding programs to develop high-yielding varieties with favorable agronomic values and better seed oil and meal quality.

1.2 Nutritional Composition

Cultivars of *B. juncea* are grown mainly for the extraction of edible oil from the seeds. Besides, they are used as spice and condiments, salad crop, and as green vegetables. *B. juncea* is also now being used as green manure or as fodder crop, and for industrial purposes due to its high erucic content in the oil (Thomas et al. 2012). Through its different edible parts, the mustard plant carries a whole range of nutraceuticals (Fig. 1). According to the USDA (USDA 2019), the seeds of Indian

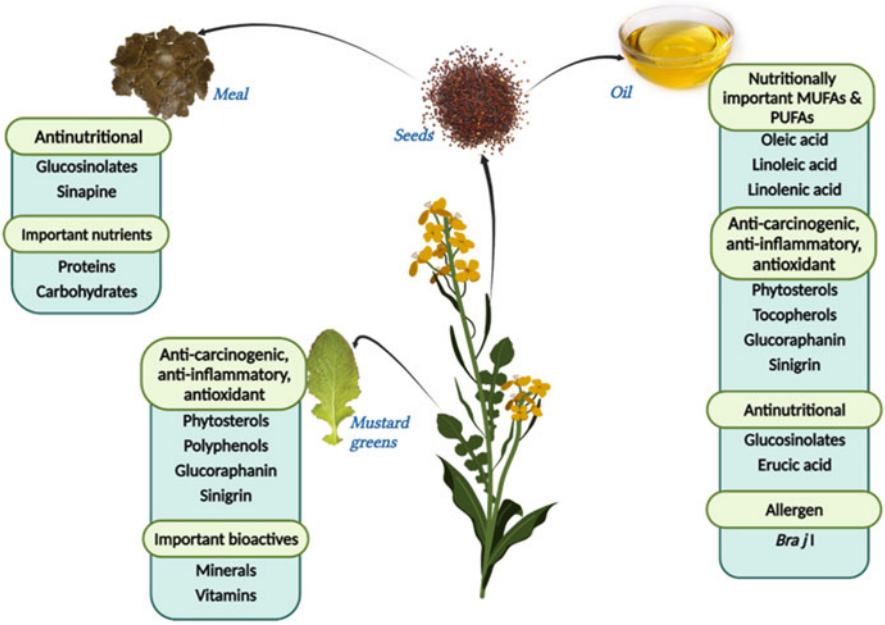


Fig. 1 Different nutraceutically important compounds present in *B. juncea*

mustard are a rich source of nutritionally vital minerals like potassium, calcium, phosphorus, and magnesium, vitamins, and dietary folate.

The calorific value of the oil of brown mustard seeds is estimated to be 541, which, although considered high, does not reduce the importance of mustard oil as a heart-healthy oil because it contains omega 3 fatty acids. *B. juncea* oil is a source of many important fatty acids like oleic acid, linoleic acid, linolenic acid, and erucic acid. However, high erucic in mustard oil is considered to be harmful and can impair myocardial conductance, cause myocardial fibrosis, and increase the blood cholesterol level (Gopalan et al. 1974; Ackman et al. 1977). *B. juncea* seeds are also rich in volatile oil (2.9%), which is optically inactive with the major constituent being allyl isothiocyanate (93–99%) formed by the hydrolysis of glucosinolate (GSL) called sinigrin an antinutritional compound.

Mustard greens, or mustard plant leaves, are a rich source of minerals, vitamins, dietary fibers, and different types of GSLs, including sinigrin and others. There are also various other types of bioactive compounds present in mustard like phenolic acids (caffeic and ferulic acid), and flavonoids (isorhamnetin, quercetin, and kaempferol). Seed meal or oil cake has a high protein, fiber, and glucosinolate content and is used mainly for feeding poultry or as a cattle feed, but also has a potential to be used as manure.

1.3 Growing Importance in the Face of Chronic Diseases and Malnutrition

Due to the presence of important fatty acids and various biologically active compounds in different plant parts of mustard, it is considered a good dietary inclusion and has been extensively used as a medicinal plant in diverse cultures since ancient times (Lietzow 2021). Major portion of mustard seeds is composed of saturated and unsaturated (monounsaturated [MUFA] and polyunsaturated [PUFA]) fatty acids. The essential fatty acids that are not synthesized in the body include oleic acid and linolenic acid (Sharafi et al. 2015), which therefore must be ingested via dietary sources (Chauhan et al. 2012). Linoleic acid has been reported to reduce the blood cholesterol level and prevent atherosclerosis (Ghafoorunissa 1994) and decrease the risk of ischemic heart disease (Rastogi et al. 2004). Introducing mustard oil in diet is therefore a good way to fulfill the body's requirement of these essential fatty acids and other important phytochemicals. In these times of growing malnutrition and prevailing diseases, introducing dietary supplements in the staple food or oil crops is an efficient approach.

Mustard has been in use in the alternative forms of medicine by different cultures around the world since early times. The ancient Greeks used mustard to cure scorpion or snakebites and to stimulate blood circulation and in the treatment of arthritis and rheumatism (Thomas et al. 2012). Mustard seeds and leaves have been used in the preparation of Indian *Ayurveda* medicine (Kapoor 1990; Khare 2004), *Yunani* medicine, and traditional Chinese (Small 2006) medicine. Prescriptions for using brown mustard seeds to treat internal and external diseases (liver and spleen enlargement) have been found in texts from Ayurveda (Rohilla et al. 2017). Skin diseases, inflammations, or rheumatism have been reportedly treated by applying mustard seeds externally (Khare 2004; Sharma and Prajapati 2016).

With the increase in focus on nutraceutical importance of spices and other food items, Indian mustard has gained further importance for its therapeutic and medicinal values in Europe and North America also. The compound isothiocyanate, which is found in abundant quantity in mustard, has emerged as a very important cancer chemopreventive phytochemical (Shin et al. 2021; Zhang 2010).

Kaplan (1989) studied the association between two major metabolic disorders, diabetes and hyperlipidemia, for their association with obesity. Based on the information available on the pharmacological activities of mustard products, they suggested their therapeutic potential against these disorders. Since Indian mustard is an easily cultivable plant and enough preclinical information is available about its medicinal uses, it can be used in breeding and germplasm improvement programs as a sustainable and affordable source of nutraceuticals that could help mitigate the malnutrition issues and also help prevent certain noncommunicable diseases and metabolic disorders.

1.4 Limitations of Conventional Breeding and Rationale for Next-Generation Breeding

The objectives of conventional *B. juncea* breeding programs have mainly focused on increasing the seed yield, oil content, and improving the quality/nutritional components of oil and meal. Although a considerable success through conventional breeding methods such as phenotypic selection, backcross breeding, and hybridization has been achieved, the presence of low variability for the oil and seed meal quality traits in the gene pool limits the success of conventional breeding programs. According to Chauhan and Singh (2004), only a few common ancestors have been used for the derivation of pure lines that were the source of most mustard varieties, and the number of donors used by the breeders was also limited. This has resulted in a narrowing of the genetic base of mustard. Various methods for genetic manipulation like mutagenesis using physical and chemical agents, hybridization, and genetic engineering are currently being used routinely to generate new variants and widen the genetic base in *Brassica* crops.

Further, the advent of molecular markers and advancements in the high-throughput omics methods have facilitated the sequencing of transcriptomes and whole genome of *B. juncea* (Yang et al. 2016; Khattak et al. 2019; Paritosh et al. 2021; Shen et al. 2021; Gao et al. 2022; Mathur et al. 2022). This has led to the identification and development of a large number of markers that can be used to dissect the genetic loci associated with various economically important traits that can now be used in precision breeding programs to improve the nutritional value of mustard oil and meal by mustard breeders.

2 Nutraceutically Important Compounds in *B. juncea*

B. juncea contains important saturated and unsaturated fatty acids, various phytosterols that are suggested to have cholesterol-decreasing properties (Duke 2002), and important secondary metabolites. The proportions of saturated and unsaturated fatty acids (MUFA and PUFA) and concentration of secondary metabolites determine the quality of mustard oil. In the seed oil, the amount of erucic acid is found to be relatively high, while the amount of glucosinolates is moderate to high.

Seed meal or oil cake left after oil extraction contains a moderate to high amounts of erucic acid and proteins (Downey and Röbbelen 1989). It also contains carbohydrates (14–15%) and fibers (10–12%) and is also an important source of phytic acid (3–6%), glucosinolates (2–3%), minerals and vitamins (1.0–1.5%), tannins (1.6–3.1%), and sinapine (1.0–1.5%). However, the presence of fibers, tannins, phytic acid, glucosinolate, and sinapine is undesirable in the seed meal, all of which lower its feed value (Chauhan et al. 2002).

Mustard greens are known to have high vitamin content, which have many health benefits (Jahangir et al. 2009; Czarnowska and Gujska 2012), fiber and mineral content (calcium, potassium), and various important bioactive compounds (Sanlier and Guler 2018).

2.1 Fatty Acids

Fatty acids are organic acids that occur in an ester or free form and contain a long-chain single carboxyl group (Chauhan et al. 2012). Saturated fatty acids do not contain a carbon–carbon double bond, for example, stearic acid (C18:0) and palmitic acid (C16:0), while unsaturated fatty acids have double bonds in their aliphatic chain, for example, erucic acid (C22:1), eicosenoic acid (C20:1), oleic acid (C18:1), linoleic acid (C18:2), and linolenic acid (C18:3). Monounsaturated fatty acids (MUFAs) contain one double bond (erucic acid and oleic acid) while polyunsaturated fatty acids (PUFAs) (linoleic and linolenic acids) contain more than one double bond in their aliphatic chain.

Two separate, genetically independent biosynthetic pathways are known for fatty acid synthesis in rapeseed-mustard (Krzymanski and Downey 1969). In one pathway, oleic acid undergoes successive desaturation and produces linoleic and then linolenic acid (Cherif et al. 1975), while, in the second pathway, oleic acid undergoes chain-lengthening process to produce eicosenoic and subsequently erucic acid (Agrawal and Stumpf 1985).

Erucic acid, with antinutritional properties, forms a major fraction (~50%) of the mustard oil (Ackman et al. 1977). Therefore, the development of *B. juncea* lines with "0" erucic acid (ZE) or <2% fraction of the seed oil has been a major objective in mustard breeding programs. Most of the *B. napus* varieties, whether pure line or hybrids, are ZE. However, ZE mustard varieties are found to have a lower seed oil content (Rout et al. 2018) than that of the normal high erucic lines (HE). Therefore, even though *B. juncea* lines with "0" erucic acid content (zero erucic mustard [ZEM]) are available for use in breeding programs (Kirk and Oram 1978), till now breeders are less successful in the development of high-oil-content-productive ZE mustard lines or varieties.

2.2 Glucosinolates (GSL)

Glucosinolates are a very important constituent of mustard plant parts, and along with erucic acid, greatly influence the quality of oil. The amount and composition of glucosinolates are different in different plant organs with the highest concentration being found in the reproductive tissues (florets, flowers, and seeds). Sinigrin, the most important glucosinolate in mustard oil, is an aliphatic compound and yields a pungent chemical called allyl isothiocyanate on hydrolysis by β-thioglucosidase enzyme (myrosinase) at the time of processing of seed for oil extraction (Yu et al. 2003). Along with allyl isothiocyanate (0.25–1.4%), glucose, potassium bisulfate, and minor amount of other volatile components are also released during hydrolysis (Thomas et al. 2012). Allyl isothiocyanate ($CH_2{=}CHCH_2N{=}C{=}S$) is a colorless liquid having a boiling point of 152 °C and acts as an antimicrobial agent (Luciano and Holley 2009). It has been shown recently as being an important phytochemical with cancer-preventive properties that also exhibits other potential health benefits (Okulicz 2010; Zhang 2010).

Some other types of glucosinolates and their breakdown products, however, may cause damage or physiological disorders. One of the harmful glucosinolates is progoitrin (2R-2-hydroxy-3-butenyl), which is a toxic indole glucosinolate and yields goitrin on breakdown (van Doorn et al. 1998). It may cause goiter or growth retardation by inhibiting the key enzyme called thyroperoxidase involved in the synthesis of thyroid hormone and is considered to be unfit for consumption by humans and animals (Rosa et al. 1997). Amongst the edible cruciferous vegetables, the highest amount of glucosinolates is reported in *B. juncea* leaves (McNaughton and Marks 2003). Cartea and Velasco (2008) have reported that goitrin can get absorbed in the human stomach by nitrosation if the water has high nitrate levels (Lüthy et al. 1984).

2.3 Vitamins and Minerals

Mustard greens are reported to be a rich source of vitamins and their precursors (vitamins A, K, C, E, folate, and carotenoids) and various important minerals (Sanlier and Guler 2018; Jahangir et al. 2009; Kim and Park 2009). According to Czarnowska and Gujska (2012), high amounts of folate in mustard may prevent cancer, vascular diseases, and neural tube defects. They are also rich in minerals important for human health like calcium (Ca), magnesium (Mg), potassium (K), chromium (Cr), iron (Fe), and phosphorus (P). Although the calcium concentration in mustard greens is relatively lower (22 mg to 150 mg/100g), a low concentration of calcium-binding compounds like oxalic and phytic acids ensures a higher bioavailability of calcium (Cartea et al. 2011) Potassium has been shown to participate in various metabolic processes (Cartea et al. 2011). *B. juncea* also has the potential to accumulate heavy metals (Del Río et al. 2004), and it thus becomes rich in important minerals like iron (Fe), zinc (Zn), and chromium (Cr), which are required in human diet.

2.4 Phenolic Compounds

Phenolic compounds present in mustard are among one of the several beneficial bioactive compounds (Sanlier and Guler 2018) that inhibit the carcinogen activity and facilitate faster detoxification of reactive oxygen species (ROS) (Morales-Lopez et al. 2017). These compounds are classified into different types according to their molecular structure: (1) simple, low molecular weight compounds, (2) single aromatic cyclic compounds, (3) tannins of large and complex nature, and (4) polyphenols derived from other sources. These compounds, based on the number and arrangement of carbon atoms, are also divided into two categories: (i) flavonoids, for example, flavonols, anthocyanidins, flavones, and flavanones; and (ii) nonflavonoids, for example, phenolic acids and stilbenes.

2.5 Agronomic and Postharvest Techniques to Improve Nutraceutical Composition

Various agronomic practices have been adopted to improve the nutrient content of food or oil crops (Kumar et al. 2020). The success of these practices depends mainly on the type and nature of crop, local environment, and the geographical location. Agronomic practices exploit the available genetic variation within the crop variety and make use of the plant's response to various types of fertilizers to improve the nutritional quality of the crop. A disadvantage of this method is the effect of chemical fertilizers on the soil and local fauna and flora.

Zinc (Zn), boron (B), potassium (K), phosphorus (P), and iron (Fe) are important minerals that have been shown to accumulate in *Brassica* plants by providing fertilizers rich in these minerals. Zn fertilizers are reported to increase the seed yield of rapeseed mustard (Grewal and Graham 1999). Nelson et al. (2016) reported that the flowering time is the crucial stage at which sufficient nutrient supply is very important to produce maximum yield.

Although glucosinolates and their derivatives are considered to be antinutritional, there are studies that have also identified the anticancerous properties of these compounds and suggested their use in crop protection and as biofumigants. Sulfur is an important constituent of glucosinolates, and in a study by Hassan et al. (2007), sulfur levels were shown to have a linear correlation with glucosinolate levels in plants. Studies in rocket salad (*Eruca sativa*) have demonstrated that seed glucosinolate levels increased on the application of fertilizers rich in sulfur and nitrogen (Singh et al. 1999). Environmental stresses have also been shown to increase the concentration of certain important compounds (l-ascorbic acid, glucosinolates, and phenolic compounds) in *Brassica* plants (Antonious et al. 2009).

Nitrogen (N) fertilizers have been shown to affect the oil quality and quantity of Indian mustard seeds. One of the objectives of oil quality improvement is to increase the concentration of oleic acid, and essential fatty acids – linoleic (ω6) and linolenic acids (ω3) – and to decrease the erucic acid concentration. According to Gao et al. (2010), high level of N increases the concentration of total saturated fatty acids and erucic acid, resulting in a decrease in oil quality.

Digestibility of seed meal is also affected by nitrogen. Plants treated with higher concentration of N show reduction in fiber content and soluble sugar levels. However, total soluble proteins do not show any significant change in N treatment (Mawlong et al. 2017). Oil pigments like β-carotene, pheophytin, and chlorophyll also showed change when the plants were treated with nitrogen (Mawlong et al. 2018).

Selenium (Se) is a micronutrient that is required in human and animal diet. Se-rich fertilizers can therefore be used to fortify mustard oil or cake to supplement the requirement of this micronutrient (White and Brown 2010). Application of Se in the form of fertilizer has also been reported to enhance the productivity of Indian mustard (Singh et al. 1980).

Some studies have shown that the fatty acid profile of *Brassicas* is also affected by the planting density and shading, wherein the erucic acid concentration was found to be reduced at high-density planting (Khan et al. 2018). Since these agronomic and postharvest methods are highly influenced by the geographical and environmental conditions of a particular crop season and are dependent upon the variability available within the species, there is a dire need to develop better and more efficient methods to improve the quality of oil and meal of *B. juncea*. Also, chemical fertilizers are harmful to the environment as they get washed away with irrigation water and contaminate the waterbodies.

Biotechnological methods are therefore a better alternative to improve oil quality traits and biofortification of mustard. These methods also allow production of new varieties and increase the overall variability within the species. Mutation breeding is another very important method to increase variability and has been used for a long time in breeding programs. Genetic modification techniques are highly effective in developing fortified food crops with high levels of health beneficial bioactive compounds. For example, the glucosinolate called glucoraphanin has many health benefits, but is not found in *B. juncea*. It will be beneficial to fortify *B. juncea* with glucoraphanin and reduce other harmful glucosinolate contents (Augustine and Bisht 2015). Using biotechnological methods, Augustine and Bisht (2015) successfully developed a *B. juncea* line (Varuna) with high glucoraphanin content.

3 Genetic Resources of Health-Related Genes

Although *B. juncea* being an amphidiploid species consists of the whole genomes of both the diploid parents (*B. rapa* and *B. nigra*), it has remained mostly unchanged since the hybridization event (Axelsson et al. 2000). Various human or natural phenomena have helped in the evolution of *B. juncea* genome and increasing the genetic variability. While natural hybridization events facilitated the genome evolution of *Brassica* species (Sharma et al. 2022), the diversity of the *B. juncea* cultivars was enriched during the domestication process by divergent selection. There are several reports using cytogenetic, genomic, and molecular marker-based investigations that facilitated the identification of basic *Brassica* crop genomes and the mode of evolution of *Brassica* and its related genera (Prakash and Hinata 1980; Hosaka et al. 1990; Song et al. 1990; Branca and Cartea 2011).

Wild races of the crop species are a very important source of genes for desirable traits and have been exploited by breeders for introgression of these genes, cytoplasmic substitutions, and construction of chromosome maps for a long time. The domestication and selection processes, although important from the plant breeding perspective, have left the gene pool of crop species very diminished over the years. Besides other factors, such as the production of F_1 hybrids, environmental disasters, agricultural practices, deforestation, and migration of population from rural to urban areas, have also added to the genetic drift. This reduction of genetic variability has resulted toward not only in the loss of wild germplasm but also in the loss of evolved landraces.

Brassica germplasm collections are mainly maintained in the form of seeds under long-term storage conditions so that the seeds could maintain viability for many years. There are many germplasm collections maintained at different places to ensure the availability of accessions for use by plant breeders and researchers. Universidad Polite´cnica of Madrid (UPM), Spain, has the one such largest repository of wild *Brassica* accessions and related genera. The National Gene Bank (NGB) in India has a diverse collection of *Brassica* germplasm (13,239) acquired from various agro-climatic zones of the country. Ninety-eight percent of the total *Brassica* collection are cultivated species while the rest are wild relatives of important *Brassica* species, out of which over 57% (7756) of all *Brassica* accessions in NGB are reported to be of *B. juncea*.

There are two important gene pools of *B. juncea* in cultivation, Indian and East European. Maximum heterosis has been recorded between the Indian × East European crosses because of the high genetic diversity between them. This variation was exploited in the development of the first commercial hybrid, DMH1, in *B. juncea* by crossing Pusa Bold and EH-2, lines from the Indian and East European gene pools, respectively (Sodhi et al. 2006).

According to two authors, Harlan and de Wet (1971), who were pioneers of the crop diversity conservation movement, the gene pool is categorized into three types (primary, secondary, and tertiary gene pools).

3.1 Primary Gene Pool

Primary gene pool is a collection of germplasm, members of which can be easily crossed, resulting in a successful sexual reproduction and fertile offspring. It includes true biological species that can be easily crossed, there is no abnormal chromosome pairing during meiosis, and genes are segregated normally. The biological species include all the spontaneously evolved wild varieties of the crop species as well as their cultivated forms (Harlan and de Wet 1971).

Each cultivated *Brassica* species, including *B. juncea* (AABB), is characterized by its own primary gene pool with several subspecies and morphological varieties, East European and Indian gene pools of *B. juncea* are one of such examples.

3.2 Secondary Gene Pool

This is a group of germplasm where crossing between the members is difficult and requires methods to overcome some biological barriers before crossing. The hybrids produced are either weak or partially sterile, and the chromosome pairing is poor. This gene pool includes non-conspecific wild relatives of the crop species and other cultivated species from the same genus, for example, *B. nigra, B. napus,* and *B. rapa*. When the genetic variability within the species is not sufficient to produce desired results, introgression from members belonging to the secondary gene pool becomes very important.

3.3 Tertiary Gene Pool

Crossing a crop variety with the members of tertiary gene pool produces anomalous, nonviable, or completely sterile hybrids. It is difficult to transfer genes for the trait of interest from a plant belonging to the tertiary gene pool to the desired crop species. Gene transfer either does not occur or requires advanced breeding or biotechnological techniques, like embryo rescue, tissue culture, etc.

The tertiary gene pool is made up of the distant wild relatives of the species, including those from different genera, for example, *Arabidopsis*. This group includes the related genera of *Brassica* species, like *Eruca, Hirschfeldia, Sinapis, Trachystoma, Enarthrocarpus,* and *Raphanus* (Hu et al. 2009; Hammer et al. 2013). Some of these genera have been used to introgress various traits of interest using somatic hybridization (Kirti et al. 1992; Müller et al. 2001; Hu et al. 2009; Hammer et al. 2013; Kumar et al. 2020).

While the variation within *B. juncea* species is limited, there are several wild relatives belonging to different genera that are considered a reservoir and a good source of novel genes of importance that can be introgressed into the cultivated varieties to improve various traits. *Crambe abyssinica* is such a source of high erucic acid genes, which is important for industrial application of the oil, while *Sinapis incana* and *Diplotaxis tenuifolia* are sources of cytoplasmic male sterility genes (Katche et al. 2019).

4 Breeding *B. juncea* Using Classical Genetics and Traditional Breeding Methods

4.1 Genetics of the Health-Related Genes

The gene controlling linolenic acid was reported as dominant by Kondra and Thomas (1975). Woods et al. (1999) found in a study that low oleic acid (C18:1) and high linoleic acid (C18:2) contents were affected by a single gene, which was observed to be under embryonic control and not the maternal parent. The gene was reported to be dominant with a small additive effect. Several authors have reported that there are two genes controlling the amount of erucic acid (C22:1) in mustard oil and both have additive effect (Tiwari 1995; Potts and Males 1999; Bhat et al. 2002; Singh and Singh 2015). The candidate gene for oleic to erucic acid elongation was reported to be *Fatty Acid Elongase 1* (*FAE1*) gene (James and Dooner 1991). Later Gupta et al. (2004) cloned and sequenced the *FAE1* gene in high as well as low erucic acid lines and identified two *FAE1* genes (*FAE1.1* and *FAE1.2*). Pandey et al. (2013) studied the erucic acid content of hybrids and their reciprocal lines and observed that the content of erucic acid in both types of hybrids was intermediate between the two parents, which suggested that maternal effect on the erucic content was absent, and it was rather under the genetic control of the developing embryo.

In case of glucosinolates, many different genes have been reported in different germplasms. In Indian gene pool lines, 6–8 genes controlling glucosinolate content

have been reported (Stringam and Thiagarajah 1995; Sodhi et al. 2002), while in case of East European gene pool, digenic control was reported (Love et al. 1990). Variety of roles played by glucosinolates in the plant system is possible due to variety of structural configurations of glucosinolates and their concentration. This variation is controlled by four major loci, *GSL-AOP*, *-ELONG*, *-OH,* and *-OX* (Kumar et al. 2020). These four loci have been shown to have epistatic effects, resulting in the generation of 14 different structural profiles of glucosinolates in plants.

4.2 Breeding Objectives: Positive and Negative Selection

Nutraceutical improvement of oilseed crops has now become an important part of research objectives. The *Brassica* breeding programs have mainly provided emphasis on increasing seed and oil yield, and improving the quality of oil (low erucic acid, high oleic acid, linoleic acid, and tocopherols) and seed meal (low glucosinolate). Since oil quality in *B. juncea* depends upon the concentration of unsaturated and saturated fatty acids, nutritional (vitamins, minerals, etc.) and antinutritional components (glucosinolate, erucic acid) present in the oil, selection for these traits is undertaken to produce biofortified mustard oil.

Positive selection for a desired trait selects for the presence or increase in the trait, like oleic acid, linoleic acid, vitamins, etc., while negative selection results in the removal or reduction of the trait phenotypes that are not desirable in the oil, like glucosinolates and erucic acid. Over 120 variants of glucosinolates are known, out of which Brassicaceae members contain one or more types in their seed or meal or greens (Fenewick et al. 1983). Thus, it is imperative to reduce glucosinolate in Indian mustard cultivars from their present high levels.

Visual selection for quality traits is ineffective, and for which marker-assisted selection becomes a useful solution. Several types of molecular markers have been identified over the years, and with the advancements in sequencing technologies, their development has become easier. This has helped tremendously in the construction of dense linkage maps and identification of quantitative trait loci (QTLs) to be used for the marker-assisted selection (MAS) of the traits of interest.

4.3 Classical Breeding Achievements: Composition and Contents

Although *B. juncea* is a mostly self-pollinated crop, cross-pollination instances of around 18% have also been recorded (Labana and Banga 1984; Bhajan et al. 1991; Abraham 1994).

The research work for the improvement of rapeseed mustard in India started at Pusa (Bihar) at the start of the twentieth century, which involved collection of different landraces and their purification (Chauhan et al. 2012). Until 1970, the only breeding methods in use were pure line and mass selection that had been used to develop several varieties, the first of which was released in 1936. In

B. juncea, the most common method used for cultivar development was pedigree selection that was responsible for the yield improvement in conventional *Brassica* breeding (Downey and Rimmer 1993). But the method is not very efficient as the environmental influence on yield components modulates the response to selection (Thurling and Depittayanan 1992).

After Sun (1943) reported the presence of heterosis in rapeseed-mustard and Ogura (1968) produced *B. juncea* lines with cytoplasmic male sterility developed by introgressing genes from *R. sativus*, hybrid production to exploit heterosis became a very important method of mustard plant breeding. The number of hybrid varieties produced therefore increased consistently after 1980. Simultaneously, mutation breeding was also pursued and has since been used successfully to improve quantitative as well as qualitative traits in *Brassicas* (Röbbelen 1990; Bhatia et al. 1999). Over 31 varieties of *B. juncea* developed by mutation breeding have been released all over the world, out of which 8 were contributed by India. A somaclone (Bio-902 or Pusa Jai Kisan) of the *B. juncea* line Varuna developed through tissue culture has also been released as a variety (Katiyar 1997).

Breeding mustard varieties with low glucosinolate content and/or low or zero erucic acid content to meet the internationally acceptable standard of oil and seed meal quality is a very important objective of mustard breeding programs in India. The first low erucic acid variety (Pusa Karishma) of Indian mustard was released in 2004, and the first double low variety (GSC 5 of gobhi sarson) that is characterized by low glucosinolate as well as low erucic acid was released in 2005. Remarkable success has been achieved in improving the oil quality by manipulating its fatty acid composition. Single mutation using ethyl methane sulfonate (EMS) for high oleic acid (70%) has also been reported in *B. juncea* (Ripley et al. 2014).

4.4 Limitations of Traditional Breeding and Rationale for Molecular Breeding

Traditional breeding methods utilize already existing genetic variation within a species. Pedigree selection, recurrent selection, backcross breeding, and hybridization, all require the presence of naturally occurring variation, which is, however, limited. Production of hybrid varieties to exploit heterosis is a very important contribution of genetics to the plant breeding programs around the world. Heterosis is generally observed to be higher in hybrids produced by crossing geographically diverse lines in comparison with the hybrids of closely related lines (Pradhan et al. 1993). However, limited variability exists within the species for oil and seed meal quality traits.

Mutagenesis has been used to introduce variations to the *B. juncea* germplasm. As the global demand of edible oil is rising and there is a need to combat malnutrition and metabolic disorders, molecular breeding has become a good alternative to traditional breeding. It allows introduction of novel genes in the mustard plant, which is not possible by traditional breeding due to biological barriers or low variability. Yao et al. (2003) transformed the *B. juncea* plants with *ADS1* gene from *Arabidopsis*, which resulted in a significant decrease in undesirable saturated fatty acids in seeds.

5 Genetic Diversity Analysis

5.1 Phenotype-Based Diversity Analysis

Phenotyping for diversity analysis is the simplest way of quantifying genetic variation and to assess the performance of the genotypes. However, phenotype-based evaluation of genetic diversity has very limited success as the relatively fewer number of traits showing polymorphism across different genotypes pose a major roadblock, while low heritability and environmental effects on the phenotypic performance are some other hurdles.

Considerable variation has been reported for various important agro-morphological traits (Khulbe et al. 2000; Pant and Singh 2001; Singh and Chowdhury 1983) among different accessions collected from various agro-climatic regions of India. Variability among different accessions can be studied by various methods; however, the first step in this process is to provide morphological characterization of germplasm for their description and classification (Smith and Smith 1989). There have been numerous studies using morphological, physiological, and molecular markers for diversity analysis in mustard (Ghosh et al. 2019; Verma et al. 2021; Chaturvedi et al. 2021).

Seeds of 20 accessions belonging to the six *Brassica* species (*B. napus*, *B. juncea*, *B. carinata*, *B. oleracea*, *B. nigra,* and *B. rapa*) including five wild relatives were utilized for analyzing oil and fatty acid composition (Sharafi et al. 2015). The fatty acids that were analyzed included palmitic, stearic, oleic, linoleic, linolenic, and erucic acid. The study observed that these fatty acids were present in the range of 89–94% for the total fatty acids in all the six species. Highest oleic acid (61%) and lowest erucic acid (1%) content were observed in *B. napus,* while in *B. juncea* the highest linolenic (20.22%) and linoleic (19.92%) acid contents were observed. Other fatty acids like oleic acid ranged from 12.46 to 16.69%, and linoleic acid content varied from 17.46 to 22.39% in the *B. juncea* varieties, while the linolenic acid content varied from 18.35 to 22.80%. Erucic acid content was also observed to be higher in *B. juncea* accessions and varied from 23.75 to 38.19%. These results clearly indicate the presence of high genetic variation in *B. juncea* for oil content and fatty acids composition.

Yadav et al. (2017) evaluated 148 accessions of *B. juncea* representing nine different agro-ecological zones of India for 15 quantitative traits. These accessions showed large variation for the trait values, wherein maximum variation was observed in case of leaf petiole length, leaf width and number, 1000 seed weight, primary branches per plant and beak length.

5.2 Molecular Markers-Assisted Assessment

The use of phenotypic markers for genetic diversity studies always has a constraint as phenotypic traits may express in one environment while may not express in another. DNA-based markers, also referred to as molecular markers, have better potential in diversity analysis as they can be utilized without any influence of the environment or plant's growth phase. Genetic diversity is measured on the basis of the genetic distance and gives us an idea about the genetic changes that could have

taken place over the time (Ravi et al. 2003). Initial attempts to use DNA-based molecular marker in *B. juncea* were taken by Jain et al. (1994) by using 500 polymorphic RAPD markers assayed among 12 Indian and 11 East European accessions of *B. juncea*. By utilizing RAPD markers, Rabbani et al. (1998) conducted genetic diversity analysis on germplasm that included accessions from Pakistan and Japan comprising of oilseed cultivars/lines as well as vegetable types. Oilseed accessions obtained from Pakistan had a low polymorphism rate, indicating these accessions had undergone selection for the same agronomic traits over time. In another study using molecular markers, AFLP loci were used by Srivastava et al. (2001) to estimate genetic diversity in 9 synthesized and 21 natural accessions of *B. juncea*. As the advancements in the field of molecular biology continued, next-generation markers like SSR, SRAP, and other marker systems were progressively used for genetic diversity analyses in *B. juncea* (Singh et al. 2021; Wu et al. 2009); however, SSR markers were the most preferred compared to other marker systems.

5.3 Relationship with Other Cultivated Species and Wild Relatives

Genetic bottleneck has occurred in *B. juncea* mainly because of domestication and polyploidy. Intensive plant breeding efforts to select desired plant particularly for low glucosinolate and low erucic acid contents has further eroded the genetic base in *B. juncea*. The breeding strategy that can be utilized to expand genetic base is through resynthesis of *B. juncea* from cultivated or wildtype diploid progenitors as the variation present in diploid species can be used to mobilize into synthetic varieties. Introgression of variation available in wild germplasm or related species can be another option. UN (1935) was the first to report resynthesis of *B. juncea* by utilizing genomic relationship triangle of *Brassica* species. In case of *B. juncea*, the genetic diversity that is present in the wild germplasm is very extensive for nuclear as well as for the cytoplasmic genes. If the alien genetic diversity can be utilized, it can be a path-breaking step in terms of widening the germplasm base of *B. juncea*. This can be achieved by facilitating the hybridization between wild relatives of *B. juncea*. However, this process is quite cumbersome due to several hybridization barriers that first need to be overcome by various means. Although different hybridization and compatibility barriers exist, still a large number of hybrids have been developed that utilize interspecific and intergeneric crossing between *B. juncea* and wild crucifers (Prakash et al. 2009). However, to get desirable results for the introgression of nuclear-encoded genes only limited attempts have been made, and that too with limited success only.

5.4 Relationship with Geographical Distribution

Vavilov (1949) suggested that the primary center of origin of *B. juncea* of the species is in Afghanistan and surrounding regions while secondary centers of origin were in

China and eastern India. These conclusions were drawn from the diversity analysis of different lines and the relationship between them. Since then, many investigations have been conducted to study diversity among collections from different geographical centers and understand the relationship between them (Gupta et al. 1991). Both morphological and molecular markers are good tools to study divergence between the varieties having different or same centers of origin. It has been shown that the extent of heterosis is maximum in the hybrids of genotypes belonging to two geographically separate gene pools (Indian and East European gene pools) (Vaughan et al. 1963; Vaughan 1977; Sodhi et al. 2006). To establish the correlations between heterosis and genetic diversity, several attempts have been made (Gupta et al. 1991; Krishna and Ghose 1992).

5.5 Extent of Genetic Diversity

Presence of genetic diversity in the germplasm is a prerequisite for many types of breeding programs. A lot of research involving diversity analyses has been conducted using cytological, phenotypic, and molecular markers. Diversity analysis has led to the identification of two distinct gene pools in Indian mustard, Indian gene pool, and East European gene pool (Vaughan et al. 1963; Vaughan 1977). These two gene pools show many trait differences as in oil content, erucic acid content, plant height, branching, flowering time, seed color, etc. (Pradhan et al. 1993). The lines belonging to these gene pools show excellent heterosis when crossed to produce hybrids, and which has already been utilized to produce commercial hybrid varieties (Sodhi et al. 2006). Other studies using molecular markers have further confirmed the availability of two gene pools in Indian mustard (Jain et al. 1994; Srivastava et al. 2001; Burton et al. 2004). Srivastava et al. (2001) used amplification fragment length polymorphism (AFLP) markers to conduct diversity analysis and found that Indian and Chinese lines formed one cluster while East European, Russian, Canadian, and Australian accessions formed another cluster.

In a study using 60 germplasm lines (27 Indian, 25 Australian, and 8 Chinese genotypes), Meena et al. (2014) investigated the extent of variability among the accessions of *B. juncea*. They used morphophysiological and quality characters to study the differences. They found that all the three types of genotypes differed from each other in either the quality traits or physiological traits. Indian genotypes had higher erucic acid content, lower oleic acid content, and higher glucosinolate content than Australian genotypes, while Chinese lines failed to set seed in Indian conditions.

6 Molecular Mapping of Health-Related Genes and QTLs

In countries like India, the major objective of various *Brassica* breeding programs is to develop cultivars with higher seed yield and high oil content, whereas in countries like Canada and members of the European Union there is a preference for oil and

meal quality (Gupta 2016). However, in India also, the focus on improving oil quality has increased in recent years and various molecular markers have been used to construct linkage maps that are then utilized to dissect QTL for the various nutraceuticals and other agronomically important traits (Pradhan et al. 2003; Ramchiary et al. 2007; Punjabi et al. 2008; Yadava et al. 2012; Dhaka et al. 2017; Panjabi et al. 2019).

6.1 Molecular Markers

Genetic markers are tools to identify genetic differences between individual organisms or species. A genetic map indicates the linear arrangement of genes on a chromosome and is developed using meiotic recombination data of a segregating population. Molecular studies using a wide array of DNA markers have been used to assemble linkage maps in many Brassica species. Various low-throughput PCR-based markers have been used over the years for linkage and QTL analysis, for example, random amplified polymorphic DNA (RAPD) (Sharma et al. 2002), restriction fragment length polymorphisms (RFLP) (Sharma et al. 1994), amplified fragment length polymorphism (AFLP), and simple sequence repeats (SSR) (Pradhan et al. 2003). Recently, advances in high-throughput sequencing and genotyping have led to the construction of very high-density maps in *B. juncea*. Next-generation sequencing technology-based platforms like genotyping by sequencing (GBS) and Illumina Infinium Brassica 90K SNP array have also been used in the map construction using doubled haploid (DH) populations (Bhayana et al. 2020; Aakanksha et al. 2021).

6.2 Mapping of QTLs for Nutraceutical Traits

Different marker systems and mapping populations have been used to map QTLs associated with fatty acids, glucosinolates, as well as tocopherols in various studies (Table 1). Cheung et al. (1998) undertook molecular mapping of seed quality traits in *B. juncea* using an F_1DH mapping population. Glucosinolate traits mapping was undertaken using DH population using AFLP markers (Mahmood et al. 2003a), in which two and three major QTLs were identified for the 3-butenyl and 2-propenyl glucosinolates, respectively. QTL mapping with F_1DH population and advanced generations (BC_4DH) using AFLP markers led to the identification of true QTL-controlling glucosinolates based on mapping and validation in DH lines (Ramchiary et al. 2007). Based on genome survey sequences (GSS) a comparative genomics approach between *B. juncea* and *B. rapa* helped in the identification of nonsynonymous single-nucleotide polymorphisms (SNPs) that were associated with the glucosinolate trait (Yang et al. 2014). Rout et al. (2015) undertook QTL mapping in DH population using AFLP, RFLP, SSR, IP, and genic SNP markers to identify glucosinolates QTLs in *B. juncea.* Ramchiary et al. (2007) used an AFLP map for the identification of loci controlling glucosinolates in F_1DH and advanced DH lines.

Table 1 Summary of QTL studies undertaken for nutraceutical traits in Indian mustard

Trait	Markers used for mapping	Mapping population	Number of QTLs	QTL/flanking markers	LOD	PVE (%)	Linkage group/ chromosome	Reference(s)
Fatty acids								
Palmitic acid (C16:0)	AFLP, RAPD	F1-derived DH	3	E4M1_4	12.4	26.1	2	Lionneton et al. (2002)
				E4M1_15	5.5	5.5	3	
				E1M2_11	7.4	14.1	6	
	AFLP	RILs	2	qpal-2.1	2.47	12.8	2	Singh et al. (2013)
				qpal-5.1	2.18	16.1	5	
Stearic acid (C18:0)	AFLP, RAPD	F1-derived DH	1	E4M1_4	6.7	15.8	2	Lionneton et al. (2002)
Oleic acid (C18:1)	AFLP, RAPD	F1-derived DH	2	E4M1_4	20.8	51.8	2	Lionneton et al. (2002)
				E1M2_6	5.5	9.5	6	
	RFLP	Reciprocal F1-derived DH	2	E_{1a}	22.17	42.9	A8	Mahmood et al. (2003b)
				E_{1b}	16.65	28	B5	
	AFLP	F1-derived DH	2	E1 M7_11	14.6	45.7	2	Lionneton et al. (2004)
				E7 M8_1	15.9	48.4	2	
	AFLP, RFLP, IP, SSR	F1-derived DH (ZE)	3	e31m47h294–e31m60h190	5.08	13.29	A9	Jagannath et al. (2011)
				e31m47h153–R71	3.13	8.82	A3	
				e31m60v138–e50t66h139	5.9	17.62	A7	

(continued)

Table 1 (continued)

Trait	Markers used for mapping	Mapping population	Number of QTLs	QTL/flanking markers	LOD	PVE (%)	Linkage group/ chromosome	Reference(s)
	AFLP, RFLP, IP, SSR	F1-derived DH (SE)	3	1g 02090b 1g 23180b	21.18	34.89	B3	Jagannath et al. (2011)
				p58m71h154 5g 08280a	2.58	2.75	A2	
				4g 15910b 4g 26840a	22.86	39.48	A8	
Linoleic acid (C18:2)	AFLP, RAPD	F1-derived DH	1	E4M1_4	14.3	41.2	2	Lionneton et al. (2002)
	AFLP	RILs	3	qlino-2.1	4.68	24.8	2	Singh et al. (2013)
				qlino-3.1	2.29	7.6	3	
				qlino-8.1	2.01	6.9	8	
	RFLP	Reciprocal F1-derived DH	2	E_{1a}	18.6	36.4	A8	Mahmood et al. (2003b)
				E_{1b}	16.02	31.9	B5	
	AFLP	F1-derived DH	2	E1 M7_10	11.4	38	2	Lionneton et al. (2004)
				E7 M8_1	12.8	41.5	2	
	RAPD	RILs	2	$OPI10_{1000}$ – $OPK12_{1000}$	2.26	12.4	9	Aggarwal et al. (2003)
				$OPD06_{600}$ – $OPA11_{400}$	2.1	10.5	17	
	AFLP, RFLP, IP, SSR	F1-derived DH (ZE)	2	e31m47h294–e31m60h190	4.92	14.5	A9	Jagannath et al. (2011)
				e31m60v138–e50t66h139	5.63	16.91	A7	

	AFLP, RFLP, IP, SSR	F1-derived DH (SE)	3	e34m62v390h–2g 20490b	2.57	3.38	A9	Jagannath et al. (2011)
				4g 20150a–1g 23180b	17.21	33.9	B3	
				4g 15910b–4g 26840a	13.38	23.6	A8	
Linolenic acid (C18:3)	AFLP, RAPD	F1-derived DH	2	E4M1 4	4.5	10.3	2	Lionneton et al. (2002)
				E6M2_1	3.7	8.1	6	
	AFLP	RILs	3	qlin-1.1	2.68	10.5	1	Singh et al. (2013)
				qlin-6.1	3.23	21.8	6	
				qlin-6.2	2.12	14.1	6	
	RFLP	Reciprocal F1-derived DH	5	E_{1a}	8.6	15.1	A8	Mahmood et al. (2003b)
				E_{1b}	7.43	14.2	B5	
				LN_2	14.02	35.4	A4	
				LN_3	4.48	7.5	A7	
				LN_4	2.52	4.2	rA5	
	AFLP	F1-derived DH	2	E6 M8_6	3.5	13.7	2	Lionneton et al. (2004)
				E1 M7_14	4	15.5	6	
	RAPD	RILs	1	$OPD19_{900}$ – $OPH02_{500}$	2.24	13.3	13	Aggarwal et al. (2003)
	RFLP	F1-derived DH	2	X107d-X52	–	32–41	7	Cheung et al. (1998)
				X8-X168	–	23–24	13	

(continued)

Table 1 (continued)

Trait	Markers used for mapping	Mapping population	Number of QTLs	QTL/flanking markers	LOD	PVE (%)	Linkage group/ chromosome	Reference(s)
	AFLP, RFLP, IP, SSR	F1-derived DH (ZE)	3	e45m34h420–p32t77v330	2.68	7.49	A3	Jagannath et al. (2011)
				e31m58v148–e44m39h131	3.82	10.94	B7	
				5g 59140b–5g 11960	4.11	11.76	A10	
	AFLP, RFLP, IP, SSR	F1-derived DH (SE)	3	1g 23440–2g 18900	3.11	8.89	A7	Jagannath et al. (2011)
				e62m33h107–1g 78560b	3.09	8.63	B1	
				1g 24310–1g 18060	4.99	14.64	A8	
Eicosenoic acid (C20:1)	AFLP, RAPD	F1-derived DH	3	E4M7_3	4.8	10.8	1	Lionneton et al. (2002)
				E3M6_4	3.7	6.9	2	
				E4M1_4	9.9	21.3	2	
	AFLP	RILs	2	qeic-2.1	3.36	22.5	2	Singh et al. (2013)
				qeic-5.1	2.02	18.2	5	
Erucic acid (C22:1)	RAPD	RILs	2	$OPF08_{1000}$ – $OPI10_{1000}$	3.74	19.5	9	Aggarwal et al. (2003)
				$OPD06_{600}$ – $OPA11_{400}$	3.06	18.7	17	
	AFLP, RAPD	F1-derived DH	3	E4M1_13	4.8	5.8	1	Lionneton et al. (2002)

				E1M7_10	4.4	4.6	2	
				E4M1_4	18.1	24	2	
	AFLP	RILs	2	qeru-2.1	2.8	20.4	2	Singh et al. (2013)
				qeru-2.2	2.07	14.8	2	
	RFLP	Reciprocal F1-derived DH	2	E_{1a}	38.26	53.7	A8	Mahmood et al. (2003b)
				E_{1b}	27.16	32.1	B5	
	RFLP	F1-derived DH	2	X140b-X61a	–	46–54	7	Cheung et al. (1998)
				X122-X193	–	17–22	4	
	AFLP	F1-derived DH	2	E1 M7_11	19.1	55.2	2	Lionneton et al. (2004)
				E7 M8_1	25.1	64.8	2	
	AFLP, RFLP	F1-derived DH	2	ea-1	20.5	60	17	Gupta et al. (2004)
				ea-2	10.2	38	3	
Glucosinolates								
Total glucosinolate	AFLP	F1-derived DH	5	E4 M1_13	4.15	10.1	1	Lionneton et al. (2004)
				E6 M8_8	2.53	5.9	7	
				E4 M1_13	2	4.7	1	
				E1 M7_2	2.48	5.8	6	
				E3 M4_3	5.77	14.6	7	
	AFLP	F1-derived DH	5	ToGsl1	2.84	3.9	J1	Ramchiary et al. (2007)
				ToGsl2	14.66	27.64	J3	

(continued)

Table 1 (continued)

Trait	Markers used for mapping	Mapping population	Number of QTLs	QTL/flanking markers	LOD	PVE (%)	Linkage group/ chromosome	Reference(s)
				ToGsl3	3	4.15	J7	
				ToGsl4	14.06	26.44	J9	
				ToGsl5	3.18	4.86	J18	
	AFLP	BC4DH	2	ToGsl2	39.48	45.46	J3	Ramchiary et al. (2007)
				ToGsl4	18.03	15.59	J9	
Aliphatic glucosinolates	RFLP	Reciprocal F1-derived DH	4	GSL-A7	2.62	6.7	A7	Mahmood et al. (2003a)
				GSL-A3	2.67	6.7	A3	
				GSL-B5	3.81	10.3	B5	
				GSL-B3	3.98	11.1	B3	
	AFLP	F1-derived DH	4	Ali1	15.33	28.67	J3	Ramchiary et al. (2007)
				Ali2	3.05	4.15	J7	
				Ali3	14.38	26.63	J9	
				Ali4	5.51	4.69	J18	
	AFLP	BC4DH	2	Ali1	38.28	44.78	J3	Ramchiary et al. (2007)
				Ali3	17.52	15.54	J9	
Total alkenyl glucosinolates	RFLP	F1-derived DH	1	X136b-X73b	–	27–46	9	Cheung et al. (1998)
Propenyl glucosinolate	RFLP	F1-derived DH	2	X212a-X16	–	33–74	16	Cheung et al. (1998)
				X37-X62	–	12–21	5	

Butenyl glucosinolate	RFLP	F1-derived DH	3	X212a-X16	–	63	16	Cheung et al. (1998)
				X136b-X73b	–	8	9	
				X176a-X129a	–	8	5	
Pentenyl glucosinolate	RFLP	F1-derived DH	3	X213-X40	–	22–47	3	Cheung et al. (1998)
				X62-X129a	–	5	5	
				X150b-X30b	–	47	1	
2-Propenyl glucosinolate	RFLP	Reciprocal F1-derived DH	3	GSL-F	3.74	10.3	Unlinked segment	Mahmood et al. (2003a)
				GSL-A2a	10.1	40.7	A2	
				GSL-A3	2.49	5.8	A3	
3-Butenyl glucosinolate	RFLP	Reciprocal F1-derived DH	3	GSL-B3	6.99	18.2	B3	Mahmood et al. (2003a)
				GSL-A2a	7.64	35.3	A2	
				GSL-A2b	2.98	22.2	A2	
Butyl glucosinolate	RFLP	Reciprocal F1-derived DH	3	GSL-B3	3.12	7.2	B3	Mahmood et al. (2003a)
				GSL-A2a	17.4	73.7	A2	
				GSL-B3	3.24	7.4	B3	
	AFLP	F1-derived DH	3	Btyl1	9.44	19.98	J3	Ramchiary et al. (2007)
				Btyl 2	3.96	6.47	J12	
				Btyl3	17.23	41.65	J16	
	AFLP	BC4DH	5	Btyl1	10.85	6.16	J3	Ramchiary et al. (2007)
				Btyl3	15.62	10.64	J16	

(continued)

Table 1 (continued)

Trait	Markers used for mapping	Mapping population	Number of QTLs	QTL/flanking markers	LOD	PVE (%)	Linkage group/ chromosome	Reference(s)
				Btyl4	74.29	73.51	J2	
				Btyl5	8.52	7.76	J9	
				Btyl6	12.97	9.17	J17	
Propyl glucosinolate	RFLP	Reciprocal F1-derived DH	2	GSL-B8	2.69	5.9	B8	Mahmood et al. (2003a)
				GSL-A2a	17	74	A2	
	AFLP	F1-derived DH	3	Prpyl1	8.36	11.42	J12	Ramchiary et al. (2007)
				Prpyl2	6.02	4.49	J12	
				Prpyl3	26.07	51.9	J16	
	AFLP	BC4DH	4	Prpyl3	21.55	9.27	J16	Ramchiary et al. (2007)
				Prpyl4	95.84	78.05	J2	
				Prpyl5	8.51	3.72	J9	
				Prpyl6	11.87	4.84	J17	
Pentyl glucosinolate	AFLP	F1-derived DH	2	Ptyl1	4.62	16.52	J3	Ramchiary et al. (2007)
				Ptyl2	3.12	3.15	J16	
	AFLP	BC4DH	4	Ptyl1	92.55	59.85	J3	Ramchiary et al. (2007)
				Ptyl3	46.12	13.53	J2	
				Ptyl4	39.17	10.88	J9	
				Ptyl5	37.27	9.8	J17	

Sinigrin	AFLP	F1-derived DH	5	E2 M2_10	3.29	8.6	12	Lionneton et al. (2004)
				E4 M1_2	3.65	10	14	
				E4 M8_8	2.44	7.7	6	
				E2 M2_10	6.13	18.4	12	
				E6 M2_13	3.44	10.4	14	
Gluconapin	AFLP	F1-derived DH	5	E2 M2_10	4.34	15.4	12	Lionneton et al. (2004)
				E4 M1_2	4	14.3	14	
				E1 M6_15	2.69	7.3	1	
				E2 M2_10	4.82	13.5	12	
				E6 M2_13	3.81	10.3	14	
Tocopherols								
α-Tocopherol	RFLP	F1-derived DH	1	WG2A11.H3–TG2B4.E1	2.5	–	3	Marwede et al. (2005)
γ-Tocopherol	RFLP	F1-derived DH	5	MG18–MG19	7.53	7.8	5	Marwede et al. (2005)
				MG21–GATA.H3	3.37		6	
				WG1G2.H1–WG3F7.H2	2.87	8.3	9	
				RP459.H1–RP1218.H1	12.1	15.6	12	
				RP1422.E1–RP1370.H1	6.34	11.4	14	
Total Tocopherol	RFLP	F1-derived DH	2	RP459.H1–RP1218.H1	4.04	22.7	12	Marwede et al. (2005)
				RP1422.E1–RP1370.H1	3.98	20.6	14	

Marwede et al. (2005) detected QTLs for tocopherols in a DH population using linkage map based on RFLP markers.

Using DH population(s) and RFLP markers, Cheung et al. (1998) and Mahmood et al. (2003b) detected QTLs that control the fatty acid profile. Lionneton et al. (2002) used a DH population and AFLP markers to map QTLs for seven constituent fatty acids of mustard oil including erucic acid. Sharma et al. (2002) used the RAPD linkage map in a RIL population to identify two QTLs for oleic acid on two different linkage groups (LGs) of *B. juncea*. Linkage map with AFLP, RFLP, SSR, IP, and genic SNP markers was also used to identify oil content QTL in two sets of populations – one segregating for erucic acid and another with zero erucic acid lines, in their oil content (Rout et al. 2018).

In *B. juncea*, the inheritance of high linoleic acid and low oleic has been found to be controlled by a single gene with dominant gene action along with minor additive effects (Potts and Males 1999; Woods et al. 1999). Kondra and Thomas (1975) investigated the fatty acid composition of *B. napus* and reported a dominant gene effect for low linolenic acid content, additive gene action for oleic and linoleic acid in one of the three crosses studied. In the other two crosses, they discovered partial dominance effect for high oleic acid and low linoleic acid content.

6.3 Association Mapping

Several SNP markers are now available with the advent of technologies in the field of genomics, and the high-throughput next-generation sequencing (NGS) platforms like Illumina, PacBio, Nanopore, etc., have further accelerated the speed of resource generation. Genome-wide association studies (GWAS) and candidate gene-based association mapping (CG-AM) are now the ideal choices to find out novel genes and marker–trait associations for health-related traits in *B. juncea* as they overcome several limitations posed by tradition mapping methods (Brachi et al. 2011). However, association mapping for health traits in *B. juncea* is mainly confined to glucosinolate traits only (Yang et al. 2021; Akhatar et al. 2020; Tandayu et al. 2022).

7 Marker-Assisted Breeding for Nutraceutical Traits

7.1 Germplasm Characterization

The role of diverse germplasm is very crucial for utilizing genetic diversity in various breeding programs. In *B. juncea*, natural variation exists for glucosinolates (Sodhi et al. 2002), erucic acid, antioxidants, and oleic acid (Sivaraman et al. 2004; Jagannath et al. 2011). However, several efforts were also made to induce variation through tissue culture (Shyam et al. 2021), ethyl methyl sulfonate (EMS)-induced mutations (Nour-Eldin et al. 2017), and using currently available technologies like CRISPR-Cas9 system (Huang et al. 2020; Assou et al. 2022). Once the germplasm

having variation for the traits are developed, they can be utilized for deriving improved varieties through marker-assisted selection and conventional breeding approaches.

7.2 Marker-Assisted Gene Introgression

The genetic loci/QTL identified in research programs need to be transferred into elite varieties having better adaptability and higher yield. Transferring the gene/QTLs based on phenotypes in a conventional breeding program is a tedious process and often involves linkage drag. Marker-assisted gene introgression is more efficient and considered more effective compared to conventional breeding as it helps in selection at early stages and can also be used for traits where phenotyping is not possible (Collard et al. 2005).

In *B. juncea*, diverse molecular markers have been deployed to map health-related (HR) genes. For example, the fatty acid elongase (*FAE*) having two homologs *FAE1.1* and *FAE1.2* that are present on the LGs A08 and B07, respectively, were tagged after sequencing and cloning the gene (Gupta et al. 2004). Gill et al. (2021) used Kompetitive allele-specific PCR (KASPar) method for tagging *FAE* gene by employing SNPs at 735 and 1476 positions in the two homologs *FAE1.1* and *FAE1.2*, respectively. Tagging of genes will assist in transferring low erucic acid trait to *B. juncea* cultivars with the objective of developing low erucic acid lines.

For glucosinolates traits in *B. juncea*, QTL studies were performed to identify molecular markers that showed close association with the regions controlling the synthesis of glucosinolates (Mahmood et al. 2003a; Ramchiary et al. 2007; Bisht et al. 2009). Pushpa et al. (2016) further used the molecular markers to validate them on a larger set of recombinant inbred lines (RIL) and showed that the markers and germplasm developed can potentially be used for marker-assisted selection of lines in *B. juncea* with low glucosinolate content. For increasing oleic acid content in *B. juncea*, Kaur et al. (2004) introgressed the desaturase suppressor genes from *B. napus* to *B. juncea* through interspecific crosses.

7.3 Gene Pyramiding

Gene pyramiding is a method to accumulate loci identified in different parents into one genotype to get the desirable trait phenotype. Gene pyramiding is possible with two approaches – through conventional breeding and by the application of biotechnological methods like recombinant DNA technology. Conventional breeding is usually time-taking; however, it can be accelerated using molecular markers. In case of wide crosses or interspecific crosses, conventional breeding approach is not possible for gene pyramiding and in such cases biotechnological means of gene pyramiding is the ideal tool although it requires great technological expertise and infrastructure.

Currently, there are no reports on gene pyramiding of health-related genes using conventional breeding methods in *B. juncea*. However, Kumar et al. (2022) have reported the pyramiding of *glyoxalase I* (*gly I*) and *γ-tocopherol methyltransferase* (*γ-TMT*) involved in tocopherol biosynthesis, which resulted in 35–51% increase in total tocopherol content in transgenic lines. α-Tocopherol, which is an active form of vitamin E and is synthesized from γ-tocopherol, has been described to reduce the risk of different forms of cancers, heart diseases, and a number of other human diseases (Bramley et al. 2000).

7.4 Limitations and Prospects of MAS and MABCB

Marker-assisted selection (MAS) and marker-assisted back cross-breeding (MABCB), besides providing a speedy introgression of genes/QTLs into related or elite germplasm, have a few inherent limitations. The biggest hurdle that limits the wide application of MAS and MABCB is the limited availability of molecular markers as markers associated with a trait identified in one germplasm have to be confirmed and validated in the unrelated germplasm before applying MAS. Also, the QTL regions identified with genome mapping are usually large and show variable expression in different backgrounds (Singh and Singh 2015). With the advancement of sequencing technologies, all six *Brassicas* have been sequenced and it has consequently resolved the issue of inadequacy of molecular markers as a large number of SNP markers are now available for different traits. However, the use and identification of SNP markers are very low in *B. juncea* mainly because of homologous sequences that are present in different subgenomes (Yang et al. 2021).

As the sequence information of all the *Brassica* species are now available, technologies like genotype by sequencing (GBS) can generate a large number of SNPs quickly and at low cost. Besides these, the recently developed 90K SNP genotyping chip has been also used in *B. juncea* that can generate a plenty of markers distributed over entire genome in one experiment (Aakanksha et al. 2021). These advancements can help in the rapid detection of markers tightly linked to the traits of interest that will accelerate speed and accuracy of MAS and MABCB.

8 Map-Based Cloning of Health-Related Genes/QTLs

Map-based cloning, also referred to as positional cloning, utilizes the association of trait to a marker whose position is known on the chromosome. Map-based cloning is an assumption-free approach as no prior information about the gene is required. The earliest attempt to clone health-related genes in *B. juncea* was made by Venkateswari et al. (1999). The study described cloning and characterization of *FATTY ACID ELONGATION1* (*BjFAE1*) gene using PCR-amplified sequences of *A. thaliana FAE* gene.

In *B. juncea*, the gene *CYP79F1*, which is an ortholog of *A. thaliana* gene *At1g16410* involved in the biosynthesis of sinigrin (an aliphatic glucosinolate),

was isolated and characterized (Sharma et al. 2016). Sinigrin on degradation produces allyl isothiocyanate, a bacteriocide with antimicrobial activities and anticancerous properties against liver and bladder invading cancers (Wang et al. 2012; Jie et al. 2014). Functional validation of *CYP79F1* was performed genetically using 95 F_2 lines along with the development of transgenic plants using a QTL-NIL J16Gsl4 line that was lacking for sinigrin content. After genetic transformation with *CYP79F1,* sinigrin biosynthesis was observed in QTL-NIL J16Gsl4.

9 Genomics-Aided Breeding for Health-Related Traits

9.1 Structural and Functional Genomics Resources Developed

Development of genetic resources is fundamental to any genomics-based breeding effort as the availability of genomic resources helps in the efficient use of molecular markers for functional characterization of genes and QTLs. In any plant species, availability of genomic resources such as DNA-based markers, genetic maps, and genome sequencing data are essential for genetic assessment and marker-assisted breeding (Srivastava et al. 2020).

The Arabidopsis Information Resource (TAIR) (Rhee et al. 2003) was initially the sole database for Brassicaceae members. As technological advancements in the sequencing methods progressed, it led to the development of more and more information in the form of EST sequences, genome sequences, and expression data on different species. Currently, vast information on structural genomics is available in the form of Brassica Database (BRAD V3.0, http://brassicadb.cn). BRAD hosts sequencing information of 26 *Brassica* species, and along with the information on syntenic loci, it also provides information on genomic variation across the different species.

Another database, "brassica.Info," was established by the Multinational Brassica Genome Project (MBGP) in 2002 and includes information of genomes, phenomes, metabolomes, transcriptomes, and proteomes. Brassica EDB is another repository that provides information on the expression of 1,01,040 genes based on RNA-seq analysis (Chao et al. 2020). These available resources have assisted in the development of molecular markers for trait improvement. However, a single comprehensive database that integrates genomics, transcriptomics, proteomics, metabolomics, and phenomics is presently not available for *Brassica* crops.

9.2 GWAS and Genomic Selection (GS)

Genome-wide association studies (GWAS) utilize linkage disequilibrium (LD), the nonrandom association of alleles of different loci present on different chromosomes or different regions of the same chromosome. Unlike biparental QTL mapping studies, GWAS utilizes variation present in a natural population and thus reduces the time that is required to develop a segregating population for linkage mapping

studies. With the availability of more and more of structural genomics resources, GWAS has emerged as the most popular option because it can utilize a wide amount of genetic variation from populations that have undergone domestication and selection.

In *B. juncea*, Yang et al. (2021) undertook GWAS to identify loci controlling glucosinolates content and reported loci on chromosomes A02 and A09 that both contain, *BjuVA09G07110.1, BjuVA02G46870.1,* and *HAG1 (MYB28)*, the major regulatory genes controlling the glucosinolates. In another study, GWAS analysis of 92 genotypes under different nitrogen conditions identified genes involved in the biosynthesis of glucosinolates, oil, and proteins (Akhatar et al. 2020). GWAS analysis by Yan et al. (2020) recognized a large number of QTLs controlling various glucosinolates and oil and protein contents under two different nitrogen (N) regimes. The candidate genes involved in the metabolism of glucosinolates, including *QGsl.ig01.1, QGsl.atg09.1,* and *QGsl.atg11.1,* were *GSH1, GSL-ALK,* and *MYB28*, which were also identified in this study. In a recent study, 158 *B. juncea* lines were used for GWAS analysis, and it identified seven loci for total GSL content, along with 8 and 19 loci for sinigrin and gluconapin, respectively (Tandayu et al. 2022). The key genes included homologs of *MYB34, MYB28, HY5, SDI2,* and *LSU2* that were located in the regions showing strong associations with the GSL traits.

Genomic selection (GS) is a promising and novel but hitherto unexplored breeding tool that has been derived as an application of GWAS analysis. GS can be used to introgress polygenic traits that are difficult to transfer through MAS or MABCB (Pant et al. 2022). GS uses a "training population" of individuals with both phenotypes and genotypes to train the prediction model to compute genomic estimated breeding values (GEBVs). Consequently, using this trained model, the GEBVs can be estimated for the untested individuals from the "candidate population" and can be used to identify selection candidates for initiating a crossing scheme or for enhanced productivity tests (Varshney et al. 2017). As the number of GWAS studies on various traits in *B. juncea* is continuously increasing, GS can serve as an efficient tool to transfer polygenic traits or QTLs to elite varieties for the improvement of nutritional and pharmaceutical characters. It is only until recently that high-density genome-wide SNP arrays have been available in *Brassica* species, allowing every individual that is phenotyped in a breeding program that could also be genotyped using SNP arrays (Clarke et al. 2016; Mason et al. 2017).

10 Recent Concepts and Strategies Developed

10.1 Gene Editing

The crop improvement practices have progressed from conventional breeding technologies like mutation, selection, and backcross breeding to the use of modern technologies like zinc finger nuclease (ZFN), transcription activator-like effector nuclease (TALEN), and clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated protein (Cas) system (CRISPR/Cas9). These new

technologies have revolutionized the crop improvement methods as the limitations posed by the conventional breeding methods can be overcome with these technologies (Nagamine and Hiroshi 2022). CRISPR/Cas9 system combined with the SNP technologies has resulted in a surge in the genome-editing practices for crop improvement. The products obtained through CRISPR/Cas9 editing or by other nucleases are classified as site-directed nuclease systems (SDN), which are further classified into SDN-1, SDN-2, and SDN-3. Recently, the Indian government has relaxed the SDN-1 and SDN-2 categories of genome-edited plants from the stringent regulatory process, and therefore these methods can now be used in the development of *B. juncea* lines with better nutritional profiles by employing methods like CRISPR/Cas9 system.

Gene editing has not been utilized extensively for improving health-related traits. There are only a few successful examples of genome editing in *B. juncea*. In *B. juncea*, Augustine et al. (2013b) used an RNAi-based suppression method to silence the transcription factor *MYB28* to develop a low GSL mustard line. Targeted modification of glucosinolate transporters (GTRs) to reduce glucosinolates in *B. juncea* by targeting induced local lesions in genome (TILLING) method has been successfully tried (Nour-Eldin et al. 2017). Kajla et al. (2017) targeted UDP-glucose:sinapate glucosyltransferase (*SGT*) and Sinapoylglucose:cholinesinapoyltransferase (*SCT*), the two enzymes involved in the synthesis of sinapine, an antinutritive compound present in *B. juncea* seeds. The study described the use of RNAi technology to silence *SGT* and *SCT* in order to reduce sinapine content in the seeds. For the modification of allergenic *Bra j I*, a seed storage protein of the 2S albumin class, gene editing using the CRISPR/Cas9-based method was used to mutate the two homologs of *Bra j I* gene (Assou et al. 2022). Apart from these few studies, there still exists a vast scope for using gene editing methods for the improvement of *B. juncea* to develop lines that are rich sources of antioxidant molecules but also have minimal levels of antinutritional compounds.

10.2 Nanotechnology

Nanotechnology is the science of studying materials at a scale of 1–100 nanometers (nm). It has emerged as a new arena of scientific research with a very broad range of applications ranging from semiconductors, textile, space industry, healthcare, and medicine to agriculture. Nanotechnology utilizes particles of the range 1–100 nm called nanoparticles and is being considered to hold promise for increasing crop productivity in the future at a large scale (King 2017). As agriculture is a traditional industry where the economic margin of profit is subtle, and the advancements and innovations in technologies using nanoparticles, the demand for increased food supplies can be attained in a sustainable manner (Kah et al. 2019). Besides increasing productivity, plant-based edible nanoparticles can be used in the health industry to treat diseases (Zhang et al. 2016). Use of mustard extract in silver nanoparticles has been shown to be hepatoprotective and can be used in acute liver injury (Hassan et al. 2020).

B. juncea being a good source of many antioxidants and health-related compounds thus has a great potential to be used in nanoparticles-based therapeutics.

11 Genetic Engineering of Nutraceutical Traits in Mustard

Mustard is highly amenable to the techniques of genetic engineering; however, the in vitro plant regeneration frequencies of different genotypes are highly variable with the Indian gene pool lines being more responsive than those belonging to the East European and Chinese groups (Fazekas et al. 1986; Yadav et al. 1991; Pental et al. 1993; Assou et al. 2022). Agrobacterium-mediated genetic transformation of hypocotyl explants is the predominantly used method for developing transgenics in *B. juncea* (Barfield and Pua 1991; Pental et al. 1993). A number of genes regulating the nutraceutical traits have been identified and functionally validated in *B. juncea* using the transgenic approach (Table 2). Recent studies have also reported the use of CRISPR/Cas9-mediated gene editing in mustard (Wang et al. 2021; Assou et al. 2022).

11.1 Fatty Acids

The nutritional quality of mustard oil is ascertained by the relative proportions of the saturated and unsaturated fatty acids. Several attempts have been made to minimize the saturated fatty acid fractions via transgenic approaches in *B. juncea*. For example, the transfer of *ADS1* gene from *Arabidopsis* to *B. juncea* resulted in a significant decrease in the concentrations of palmitic and stearic acids, along with an enhancement in the oleic acid fraction (Yao et al. 2003). *B. juncea* lines contain a high amount of erucic acid (~50%), which is a long-chain fatty acid and an antinutritional component (Ackman et al. 1977). *FAE1* gene encodes the first enzyme in the erucic acid biosynthesis pathway, β-ketoacyl-CoA synthase (KCS) (James et al. 1995). The knockdown of *FAE1* gene in *B. juncea* causes a significant reduction in the erucic acid fractions (Kanrar et al. 2006). Sinha et al. (2007) overexpressed *FatB* from *Diploknema (Madhuca) butyracea* in *B. juncea,* which was accompanied with the silencing of endogenous *FAE1* gene using hairpin-RNA. This resulted in a 64–82% reduction in erucic acid in transgenic lines with a 4–13% enhancement in the oil content. A suppression of *FAD2*, which encodes a crucial enzyme in the biosynthesis of PUFA, has been shown to reduce α-linolenic acid content accompanied with an increase in the oleic acid content (Stoutjesdijk et al. 2000; Sivaraman et al. 2004). The PUFA (linoleic and linolenic acids) do not pose a direct heath risk to humans; however, oils with high concentrations of C18:2 and C18:3 oxidize and deteriorate rapidly, which lends them unfit for consumption (Röbbelen and Nitsch 1975). Another PUFA – γ-linolenic acid (GLA) – has been shown to possess therapeutic properties and is used as a dietary supplement (Horrobin 1992). Mustard oil lacks GLA due to the absence of *Δ6-desaturase (d6D)*, and therefore the transgenic lines of mustard with high levels of GLA and enhanced nutritional quality were engineered via the transfer of *d6D* from heterologous systems (Hong et al. 2002; Das et al. 2006).

Table 2 A summary of genetic engineering of nutraceutical traits in mustard

Trait	Added nutraceutical value	Gene	*B. juncea* genotype	Approach for genetic engineering	Reference(s)
Fatty acids	Low saturated fatty acids and high oleic acid	*ADS1* from *Arabidopsis*	J96D-4830	Transgene expression	Yao et al. (2003)
	High oleic acid	*FAD2*	815-1-6-2	Co-suppression	Stoutjesdijk et al. (2000)
	High oleic acid and low linolenic acid	*FAD2*	VH486	Antisense RNA	Sivaraman et al. (2004)
	Improved oil content	*DGAT1*	RLM198	Transgene expression	Savadi et al. (2015)
	Increase in γ-linolenic acid	*d6D* from *Pythium irregulare*	1424	Transgene expression	Hong et al. (2002)
	Increase in γ-linolenic acid	*d6D* from *Synechocystis* PCC6803	B85	Transgene expression	Das et al. (2006)
	High oleic, PUFA, and low erucic acid	*ACP* from *Azospirillum brasilense*	PCR7	Transgene expression	Jha et al. (2007)
	High oleic, PUFA, and low erucic acid	*FatB* from *Diploknema (Madhuca) butyracea* and *FAE1* from *B. juncea*	PCR7	*FatB* overexpression; silencing of endogenous *FAE1*	Sinha et al. (2007)
	Low erucic acid	*BjFAE1*	RLM-198	Knockdown	Kanrar et al. (2006)
Glucosinolates	Low aliphatic glucosinolates	*BjuMYB28*	Varuna	RNAi and Antisense RNA	Augustine et al. (2013b), Augustine and Bisht (2019)
	High glucoraphanin	GSL-ALK	Varuna	Gene silencing	Augustine and Bisht (2015)
	High sinigrin	*CYP79F1*	QTL-NIL *J16Gsl4*	Transgene expression	Sharma et al. (2016)
Others	Low sinapine	SGT and SCT	Varuna	Gene silencing	Kajla et al. (2017)

(continued)

Table 2 (continued)

Trait	Added nutraceutical value	Gene	*B. juncea* genotype	Approach for genetic engineering	Reference(s)
	High α-tocopherol (vitamin E)	*γ-TMT*	Varuna	Transgene overexpression	Yusuf and Sarin (2007)
	Reduced allergen Bra j I	*Bra j IA* and *Bra j IB*	Terratop and CR2664	Gene editing (CRISPR/Cas9)	Assou et al. (2022)
	Immunity against anthrax	PA gene from *Bacillus anthracis*	Varuna	Transgene expression	Gorantala et al. (2014)

11.2 Glucosinolates

Glucosinolates are secondary metabolites and a characteristic of the family Brassicaceae. The hydrolytic products of some glucosinolates impart nutraceutical value to mustard whereas some of them have antinutritional properties. Therefore, the reduction of nonbeneficial glucosinolates and improvement of those with therapeutic potential are the key targets of breeding programs for the genetic improvement of oilseed mustard. Transgenic approaches have used RNAi to target *BjMYB28*, a key transcription factor controlling the biosynthesis of aliphatic glucosinolates (Augustine et al. 2013a, b; Augustine and Bisht 2019). The transgenic line with the most significant suppression of the gene showed a striking 89% reduction in aliphatic fraction of seed glucosinolates without altering the levels of desirable nonaliphatic glucosinolates (Augustine et al. 2013b). However, it also showed an 80–90% decrease in leaf glucosinolate content. The leaf glucosinolates play an important role in plant defense responses (Hopkins et al. 2009). In another study, Augustine and Bisht (2019) used seed-specific *FAE1* and napin promoters in RNAi and antisense constructs to knockdown *BjMYB28*. They observed no reduction in total glucosinolates in the transgenic lines driven by napin promoters. The transgenic lines with the *FAE1* promoter showed reduction in total glucosinolates but also presented altered levels of other seed quality traits. The study concluded that the use of RNAi constructs with the native promoter could be the strategy of choice for the silencing of *BjMYB28*. Nour-Eldin et al. (2017) reported a seed-specific reduction of aliphatic glucosinolates in *B. juncea* by mutagenesis (TILLING) of 4 of the 12 *GTR* orthologs in *B. juncea* and observed a 60–70% reduction in the seed glucosinolate levels.

Glucoraphanin is another major nutraceutical compound in mustard. Its hydrolytic derivative, sulphoraphane, has anticancer properties (Fahey et al. 1997). Therefore, the biofortification of glucoraphanin is a major objective for the nutritional enhancement of oilseed mustard. The knockdown of *GSL-ALK* by Augustine and Bisht (2015) resulted in an enhancement in the amount of glucoraphanin (up to 43.11 μmol g^{-1} of dry weight) in *B. juncea*. The transgenic lines also showed reduced concentrations of gluconapin, sinigrin, and total glucosinolates. Sinigrin, which is the final product of 3C glucosinolate pathway, upon degradation produces allyl isothiocyanate that has been reported to have antimicrobial and anticancer activities (Okulicz 2010; Bhattacharya et al. 2010; Jie et al. 2014). Therefore, an increase in the levels of sinigrin is desirable for improvement of nutraceutical properties of mustard seed. Sharma et al. (2016) identified *CYP79F1* as a candidate gene for the regulation of sinigrin biosynthesis. The study showed that the transformation of wildtype *CYP79F1* gene in *B. juncea* QTL-NIL *J16Gsl4* line lacking sinigrin resulted in the biosynthesis of sinigrin in the transgenic lines.

11.3 Other Nutraceutical Compounds

The mustard seeds also contain an antinutritive phenolic compound – sinapine – which accumulates in the embryo during seed filling stages and reduces the

digestibility of *Brassica* seed meal (Huang et al. 2008). The RNAi-mediated knockdown of *SGT* and *SCT* genes, encoding enzymes crucial for final steps of sinapine biosynthetic pathway, resulted in up to 67% decrease in sinapine levels in the transgenic mustard lines (Kajla et al. 2017). These results are crucial for the improvement of *B. juncea* seed meal. Transgenic approaches have also been used for the biofortification of mustard with the components imparting high nutritive value. Yusuf and Sarin (2007) overexpressed the *Arabidopsis γ-TMT* gene in *B. juncea*. This resulted in an enhancement in the α-tocopherol levels in the seeds of the transgenic lines. α-Tocopherol/vitamin E is an essential fat-soluble vitamin with antioxidant and therapeutic properties (Bramley et al. 2000).

Another objective of mustard improvement programs is the reduction of allergens. The brown mustard seeds contain an allergen compound, Bra j I, which is a major storage protein. In a recent study, Assou et al. (2022) inactivated the *Bra j I* gene in brown mustard using CRISPR/Cas9 system and developed Bra j I-free *B. juncea* lines. The study is a big move toward the development of allergen-free mustard; however, extensive efforts are still needed to identify and characterize other compounds with allergenic properties in mustard.

Oilseed crops are preferred for the development of edible vaccines as they involve only a few processing steps (Daniell et al. 2001). Gorantala et al. (2014) expressed an antigenic protein from *Bacillus anthracis* – protective antigen (PA) to develop oral anthrax vaccine in *B. juncea*. The study demonstrated the development of systemic and mucosal immune response in mice on oral ingestion of purified PA and highlighted the scope of use of oilseed mustard for the production of biopharmaceuticals.

12 Resources for Translational Genomics in *B. juncea*

The model plant system *Arabidopsis* is the most extensively studied member of the family Brassicaceae and was also the first plant genome to be sequenced (The Arabidopsis Genome Initiative 2000). The TAIR database (https://www.arabidopsis.org/) provides extensive information on the genetics and genomics of *Arabidopsis* (Rhee et al. 2003). Cytological and genetic mapping evidence showed the presence of syntenic regions defined by gene collinearity between *A. thaliana* and the diploid *Brassica* species – *B. rapa* (AA), *B. nigra* (BB), and *B. oleracea* (CC) (Lysak et al. 2005; Parkin et al. 2005). The genome triplication was followed by a reduction in chromosome numbers and genome reshuffling in the three paleohexaploid genomes A, B, and C (Schranz et al. 2006; Mun et al. 2009). The genome assemblies of *B. rapa* (Wang et al. 2011) and *B. oleracea* (Liu et al. 2014) have provided further credence to these hypotheses and have revealed that the three diploid genomes underwent gene fractionation. Subsequent hybridization between the three diploid species led to the evolution of the three allopolyploid species of U's triangle (UN 1935).

The genomes of all the species of the U's triangle have now been sequenced, and the resulting data on genomic and genic sequences is freely available in public databases, including NCBI, *Brassica*DB (http://brassica.nbi.ac.uk/BrassicaDB/),

Brassica.info (http://www.brassica.info/), and the *Brassica* database (BRAD; http://brassicadb.cn). BRAD also includes the genome assembly of *B. juncea* var. Tumida (Yang et al. 2016) along with 25 other species of Brassicaceae and provides their syntenic relationships to the genes in *Arabidopsis* (Chen et al. 2022). Since there is a close phylogenetic relationship between the species of Brassicaceae, these resources could be useful for the identification of candidate genes for different agronomic and nutritional traits using comparative genomics. The recently published, highly contiguous genome assembly of an Indian oleiferous line – Varuna (Paritosh et al. 2021) – would be of immense value to dissect the QTL for different agronomic and nutritional traits mapped previously in *B. juncea* (Ramchiary et al. 2007; Panjabi-Massand et al. 2010; Yadava et al. 2012; Padmaja et al. 2014; Dhaka et al. 2017; Rout et al. 2018). Recently, Kang et al. (2021) have also published a contiguous assembly of a yellow-seeded *B. juncea* var. Sichuan Yellow (SY) and also resequenced 480 global accessions of *B. juncea,* which included accessions from previously defined Indian and East European gene pools (Pradhan et al. 1993; Srivastava et al. 2001) and exotic accessions predominantly from East Asia. These studies provide scope for the use of a wider germplasm to study allelic diversity and would considerably accelerate the translational research in mustard.

Seed is the economically most important product and is the primary source of nutraceutical compounds in oilseed mustard. Recently, Gao et al. (2022) reported a spatiotemporal transcriptome atlas of seed coat and embryo development of six *Brassica* species of U's triangle, including *B. juncea* (AC Vulcan-J). The study revealed differences in the accumulation of storage reserves and fatty acid metabolism among these species, and is of immense value toward understanding the transcriptional dynamics underlying different seed traits. A number of studies have also undertaken a comparative transcriptome profiling of seeds from diverse *B. juncea* lines (Khattak et al. 2019; Shen et al. 2021; Mathur et al. 2022) and provided detailed insights into understanding the biology of these traits.

Additionally, the biosynthesis and degradation pathways of most nutraceutical compounds like fatty acids, glucosinolates, flavonoids, and phenolic compounds have been well characterized in the model plant systems and are freely accessible in databases, including KEGG (Moriya et al. 2007), AraCyc (Mueller et al. 2003), PlantCyc (Grafahrend-Belau et al. 2012), and ARALIP (Li-Beisson et al. 2013), among others, and will be useful in the translational research directed at the improvement of these traits.

13 Social, Political, and Regulatory Issues in the Improvement of Nutraceutical Traits

The transgenic approaches have been successful in the genetic engineering of a number of nutraceutical traits (Table 1), especially for traits for which very low/no variability is reported in the *B. juncea* germplasm limiting the use of conventional breeding approaches, for example, enhancement of GLA and glucoraphanins (Hong et al. 2002; Das et al. 2006; Augustine and Bisht 2015). Additionally, the transgenic

approaches are more efficient, faster, and most importantly avoid the problem of linkage drag. However, their use in edible crops involves a number of ethical, socioeconomic, and regulatory concerns. The technique has been opposed by environmentalists across the world on the grounds of biosafety issues. The genetically modified crops are subjected to regulated field trials and strict biosafety checks before their commercial release. This limits the scope of improvement of nutraceutical compounds in mustard using these approaches. However, a number of countries including India have relaxed the regulatory norms for crops developed using gene-editing approaches like the novel CRISPR/Cas9 system. This would prove to be beneficial for the improvement of traits where gene knockout/allelic conversions/expression changes by modification of promoter sequences result in an altered phenotype. Recent reports have successfully used CRISPR/Cas9 approach to improve nutraceutical (Assou et al. 2022) and yield traits (Wang et al. 2021) in *B. juncea,* which should be seen as a major development in the genetic improvement of mustard.

14 Conclusions and Future Outlook

The enhancement of oil quality and nutritional composition are the major breeding objectives for the improvement of oilseed mustard. Since mustard is the predominant source of edible oil in India, more research initiatives are directed toward the improvement of nutritional value of oil in the Indian mustard lines by the reduction of antinutritive components – erucic acid, linoleic acid, α-linolenic acid, aliphatic seed glucosinolates, and sinapine, accompanied with the biofortification of nutraceuticals – oleic acid, glucoraphanin, sinigrin, GLA, tocopherols, among others, to enhance the nutritive value of mustard oil. Extensive research has been undertaken for molecular mapping of these traits in mustard; however, the QTL mapping studies have predominantly exploited the genetic diversity between the Indian and East European gene pools only. The availability of genome sequence information of a large number of global accessions of *B. juncea* could now be exploited in conjunction with the high-throughput genotyping platforms and large-scale phenotyping experiments can be undertaken for studying the extent of allelic diversity and identification of novel variants for these traits in a wider germplasm. Furthermore, the availability of genome sequences of the East European and Indian gene pool lines of *B. juncea* could also be used to identify the genes underlying the QTL identified for these traits using candidate gene approaches. Moreover, the availability of genome and comparative seed transcriptome data of all the species of U's triangle will tremendously aid the translation of information available in these crops to enhance the nutraceutical traits in *B. juncea* using comparative genomics. Further, the recent advancements in the optimization of the protocols for gene editing using CRISPR/Cas9 in both East European and Indian *B. juncea* lines appear promising in expediting the targeted improvement of quality traits in mustard with relatively fewer biosafety concerns. Therefore, there is a constant need to identify suitable target genes to leverage the immense potential of these resources for the

improvement of *B. juncea*. Finally, studies need to be undertaken to identify and characterize other compounds with a potential nutraceutical value in *B. juncea* to enhance its food and seed meal quality.

References

Aakanksha, Yadava SK, Yadav BG et al (2021) Genetic analysis of Heterosis for yield influencing traits in *Brassica juncea* using a doubled haploid population and its backcross progenies. Front Plant Sci 12. https://doi.org/10.3389/fpls.2021.721631

Abraham V (1994) Rate of out crossing in Indian mustard (*Brassica juncea* L.). Eucarpia Cruciferae News Lett 16:69–70

Ackman RG, Eaton CA, Sipos JC, Loew FM, Hancock D (1977) A comparison of fatty acids from high levels of docosenoic acids of rapeseed oil (erucic acid) and of partially hydrogenated fish oil (primarily cetoleic acid) in a non-human primate species in a short-term exploratory study. Bibl Nutr Dieta 25:170–185. https://doi.org/10.1159/000400508

Aggarwal RAK, Sharma R, Kumar R et al (2003) Molecular mapping of loci affecting the contents of three major fatty acids in Indian mustard (*Brassica juncea* L). J Plant Biochem Biotechnol 12: 31–137. https://doi.org/10.1007/BF03263173

Agrawal VP, Stumpf PK (1985) Elongation systems involved in the biosynthesis of erucic acid from oleic acid in developing *Brassica juncea* seeds. Lipids 20:361–366

Akhatar J, Singh MP, Sharma A et al (2020) Association mapping of seed quality traits under varying conditions of nitrogen application in *Brassica juncea* L. Czern & Coss. Front Genet 11: 744. https://doi.org/10.3389/fgene.2020.00744

Antonious GF, Bomford M, Vincelli P (2009) Screening *Brassica* species for glucosinolate content. J Environ Sci Health B 44(3):311–316

Assou J, Zhang D, Roth KD, Steinke S, Hust M et al (2022) Removing the major allergen *Bra j* I from brown mustard (*Brassica juncea*) by CRISPR/Cas9. Plant J 109:649–663. https://doi.org/10.1111/tpj.15584

Augustine R, Bisht NC (2015) Biofortification of oilseed *Brassica juncea* with the anti-cancer compound glucoraphanin by suppressing GSL-ALK gene family. Sci Rep 5:1–12. https://doi.org/10.1038/srep18005

Augustine R, Bisht NC (2019) Targeted silencing of genes in polyploids: lessons learned from *Brassica juncea*-glucosinolate system. Plant Cell Rep 38:51–57. https://doi.org/10.1007/s00299-018-2348-8

Augustine R, Majee M, Gershenzon J, Bisht NC (2013a) Four genes encoding *MYB28*, a major transcriptional regulator of the aliphatic glucosinolate pathway, are differentially expressed in the allopolyploid *Brassica juncea*. J Exp Bot 64:4907–4921. https://doi.org/10.1093/jxb/ert280

Augustine R, Mukhopadhyay A, Bisht NC (2013b) Targeted silencing of *BjMYB28* transcription factor gene directs development of low glucosinolate lines in oilseed *Brassica juncea*. Plant Biotechnol J 11:855–866. https://doi.org/10.1111/pbi.12078

Axelsson T, Bowman CM, Sharpe AG et al (2000) Amphidiploid *Brassica juncea* contains conserved progenitor genomes. Genome 43:679–688. https://doi.org/10.1139/g00-026

Barfield DG, Pua EC (1991) Gene transfer in plants of *Brassica juncea* using *Agrobacterium tumefaciens*-mediated transformation. Plant Cell Rep 10:308–314. https://doi.org/10.1007/BF00193148

Bhajan R, Chauhan YS, Kumar K (1991) Natural cross pollination in Indian mustard. Cruciferae Newslett 14:24–25

Bhat MA, Gupta ML, Banga SK, Raheja RK, Banga SS (2002) Erucic acid heredity in *B. juncea* – some additional information. Plant Breed 121:456–458

Bhatia CR, Nichterlein K, Maluszynski M (1999) Oilseed cultivars development from induced mutations and mutation altering fatty acid composition. Mut Breed 11:1–36

Bhattacharya A, Li Y, Wade KL, Paonessa JD, Fahey JW et al (2010) Allyl isothiocyanate-rich mustard seed powder inhibits bladder cancer growth and muscle invasion. Carcinogenesis 31: 2105–2110. https://doi.org/10.1093/carcin/bgq202

Bhayana L, Paritosh K, Arora H et al (2020) A mapped locus on LG A6 of *Brassica juncea* line Tumida conferring resistance to white rust contains a CNL type R gene. Front Plant Sci 10:1690. https://doi.org/10.3389/fpls.2019.01690

Bisht NC, Gupta V, Ramchiary N et al (2009) Fine mapping of loci involved with glucosinolate biosynthesis in oilseed mustard (*Brassica juncea*) using genomic information from allied species. Theor Appl Genet 118:413–421. https://doi.org/10.1007/s00122-008-0907-z

Brachi B, Morris GP, Borevitz JO (2011) Genome-wide association studies in plants: the missing heritability is in the field. Genome Biol 12:232. https://doi.org/10.1186/gb-2011-12-10-232

Bramley PM, Elmadfa I, Kafatos A, Kelly FJ, Manios Y et al (2000) Vitamin E. J Sci Food Agric 80:913–938

Branca F, Cartea E (2011) Brassica. In: Kole C (ed) Wild crop relatives: genomic and breeding resources, oilseeds. Springer, Berlin/Heidelberg, pp 17–36

Burton W, Ripley V, Potts D, Salisbury P (2004) Assessment of genetic diversity in selected breeding lines and cultivars of canola quality *Brassica juncea* and their implications for canola breeding. Euphytica 136:181–192

Cartea ME, Velasco P (2008) Glucosinolates in Brassica foods: bioavailability in food and significance for human health. Phytochem Rev 7:213–229. https://doi.org/10.1007/s11101-007-9072-2

Cartea ME, Lema M, Francisco M, Velasco P (2011) Basic information on vegetable Brassica crops. In: Genetics, genomics and breeding of vegetable Brassicas, pp 1–33

Chao H, Li T, Luo C et al (2020) BrassicaEDB: a gene expression database for Brassica crops. Int J Mol Sci 21:5831. https://doi.org/10.3390/ijms21165831

Chaturvedi VD, Maurya KN, Upadhyay DK, Singh PK, Chaubey S (2021) Genetic diversity for yield and its contributing components in Indian mustard [*Brassica juncea* (L.) Czern & Coss.]. J Pharm Innov 10:748–754

Chauhan JS, Singh NB (2004) Breeding approaches in rapeseed mustard varietal improvement. (in) rapeseed- mustard research in India. In: Singh NB, Kumar A (eds) . National Research Center on Rapeseed-Mustard, Bharatpur, pp 51–64

Chauhan JS, Tyagi MK, Kumar PR, Tyagi P, Singh M, Kumar S (2002) Breeding for oil and seed meal quality in rapeseed mustard in India – a review. Agric Rev 23(2):71–92

Chauhan JS, Singh KH, Singh VV, Kumar S (2012) Hundred years of rapeseed-mustard breeding in India: accomplishments and future strategies. Indian J Agric Sci 81(12):1093–1109

Chen H, Wang T, He X, Cai X, Lin R et al (2022) BRAD V3.0: an upgraded Brassicaceae database. Nucleic Acids Res 50:D1432–D1441. https://doi.org/10.1093/NAR/GKAB1057

Cherif A, Dubacq J, Mache R et al (1975) Biosynthesis of α-linolenic acid by desaturation of oleic and linoleic acids in several organs of higher and lower plants and in algae. Phytochemistry 14: 703–706. https://doi.org/10.1016/0031-9422(75)83018-4

Cheung WY, Landry BS, Raney P, Rakow GFW (1998) Molecular mapping of seed quality traits in *Brassica juncea* L.Czern. and Coss. Acta Hortic 459:139–147

Clarke WE, Higgins EE, Plieske J et al (2016) A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. Theor Appl Genet 129:1887–1899. https://doi.org/10.1007/s00122-016-2746-7

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142:169–196. https://doi.org/10.1007/s10681-005-1681-5

Czarnowska M, Gujska E (2012) Effect of freezing technology and storage conditions on folate content in selected vegetables. Plant Foods Hum Nutr 67:401–406

Daniell H, Streatfield SJ, Wycoff K (2001) Medical molecular farming: production of antibodies, biopharmaceuticals and edible vaccines in plants. Trends Plant Sci 6:219–226

Das B, Goswami L, Ray S et al (2006) Agrobacterium-mediated transformation of *Brassica juncea* with a cyanobacterial (Synechocystis PCC6803) delta-6 desaturase gene leads to production of gamma-linolenic acid. Plant Cell Tissue Organ Cult 86:219–231. https://doi.org/10.1007/s11240-006-9111-5

Del Río M, Font R, De Haro A (2004) Heavy metal uptake by Brassica species growing in the polluted soils of Aznalcóllar (Southern Spain). Fresenius Environ Bull 13:1439–1443

Dhaka N, Rout K, Yadava SK, Sodhi YS, Gupta V et al (2017) Genetic dissection of seed weight by QTL analysis and detection of allelic variation in Indian and east European gene pool lines of *Brassica juncea*. Theor Appl Genet 130:293–307. https://doi.org/10.1007/s00122-016-2811-2

Downey RK, Rimmer SR (1993) Agronomic improvement in oilseed Brassicas. Adv Agron 50: 1–65

Downey RK, Röbbelen G (1989) Brassica species. In: Röbbelen G, Downey RK, Ashri A (eds) Oil crops of the world. McGraw-Hill, New York, pp 339–362

Duke JA (2002) Handbook of medicinal herbs, 2nd edn. CRC Press, Boca Raton, FL, p 516e517

Fahey JW, Zhang Y, Talalay P (1997) Broccoli sprouts: An exceptionally rich source of inducers of enzymes that protect against chemical carcinogens. Proc Natl Acad Sci USA 94:10367–10372. https://doi.org/10.1073/pnas.94.19.10367

Fazekas GA, Sedmach PA, Palmer MV (1986) Genetic and environmental effects on in vitro shoot regeneration from cotyledon explants of *Brassica juncea*. Plant Cell Tissue Organ Cult 6:177–180. https://doi.org/10.1007/BF00180802

Fenewick GR, Heaney RK, Mullin J (1983) Glucosinolates and their breakdown products in food and food plants. CRC Crit Rev Food Sci Nutr 18:123–201

Gao J, Thelen KD, Min DH, Smith S, Hao X, Gehl R (2010) Effects of manure and fertilizer applications on canola oil content and fatty acid composition. Agron J 102(2):790–797

Gao P, Quilichini TD, Yang H, Li Q, Nilsen KT et al (2022) Evolutionary divergence in embryo and seed coat development of U's Triangle Brassica species illustrated by a spatiotemporal transcriptome atlas. New Phytol 233:30–51. https://doi.org/10.1111/nph.17759

Ghafoorunissa G (1994) Dietary fats/oils and heart diseases. In: Prasad MVR (ed) Sustainability in oil seeds. Indian Society of Oil Seed Research, Hyderabad, pp 486–490

Ghosh S, Mazumder M, Mondal B et al (2019) Morphological and SSR marker-based genetic diversity analysis of Indian mustard (*Brassica juncea* L.) differing in *Alternaria brassicicola* tolerance. v. 12

Gill KS, Kaur G, Kaur G, Kaur J, Kaur Sra S, Kaur K, Gurpreet K, Sharma M, Bansal M, Chhuneja P, Banga SS (2021) Development and validation of Kompetitive Allele-specific PCR assays for Erucic acid content in Indian Mustard [*Brassica juncea* (L.) Czern and Coss.]. Front Plant Sci 12:738805

Gopalan C, Krishnamurthi D, Shenolikar IS, Krishnamachari KAVR (1974) Myocardial changes in monkeys fed mustard oil. Ann Nutr Metab 16:352–365. https://doi.org/10.1159/000175508

Gorantala J, Grover S, Rahi A, Chaudhary P, Rajwanshi R et al (2014) Generation of protective immune response against anthrax by oral immunization with protective antigen plant-based vaccine. J Biotechnol 176:1–10. https://doi.org/10.1016/j.jbiotec.2014.01.033

Grafahrend-Belau E, Junker BH, Falk S (2012) Plant metabolic pathways: databases and pipeline for stoichiometric analysis. In: Seed development: OMICS technologies toward improvement of seed quality and crop yield: OMICS in seed biology. Springer, Dordrecht, pp 345–366

Grewal HS, Graham RD (1999) Residual effects of subsoil zinc and oilseed rape genotype on the grain yield and distribution of zinc in wheat. Plant Soil 207(1):29–36

Gupta SK (2016) Brassicas. In: Gupta SK (ed) Breeding oilseed crops for sustainable production. Opportunities and constraints. Academic, Cambridge, pp 33–53

Gupta VP, Sekhon MS, Satija DR (1991) Studies on genetic diversity, heterosis and combining ability in Indian mustard. Indian J Genet 51:448–453

Gupta V, Mukhopadhyay A, Arumugam N et al (2004) Molecular tagging of erucic acid trait in oilseed mustard (*Brassica juncea*) by QTL mapping and single nucleotide polymorphisms in FAE1 gene. Theor Appl Genet 108:743–749. https://doi.org/10.1007/s00122-003-1481-z

Hammer K, Gladis T, Laghetti G, Pignone D (2013) The wild and the grown-remarks on *Brassica*. Int J Agric Sci 3(6):453–480

Harlan JR, De Wet JMJ (1971) Toward a rational classification of cultivated plants. Taxon 20: 509–517

Hassan FU, Manaf A, Qadir G, Basra SMA (2007) Effects of sulphur on seed yield, oil, protein and glucosinolates of canola cultivars. Int J Agric Biol 9:504–508

Hassan SA, el Hagrassi AM, Hammam O et al (2020) *Brassica juncea* L. (Mustard) extract silver nanoparticles and knocking off oxidative stress, proinflammatory cytokine and reverse DNA genotoxicity. Biomolecules 10:1650. https://doi.org/10.3390/biom10121650

Hong H, Datla N, Reed DW, Covello PS, MacKenzie SL et al (2002) High-level production of γ-Linolenic acid in *Brassica juncea* using a Δ6 desaturase from *Pythium irregulare*. Plant Physiol 129:354. https://doi.org/10.1104/PP.001495

Hopkins RJ, Van Dam NM, Van Loon JJA (2009) Role of glucosinolates in insect-plant relationships and multitrophic interactions. Annu Rev Entomol 54:57–83. https://doi.org/10.1146/annurev.ento.54.110807.090623

Horrobin DF (1992) Nutritional and medical importance of gamma-linolenic acid. Prog Lipid Res 31:163–194

Hosaka K, Kianian SF, McGrath JM, Quiros CF (1990) Development and chromosomal localization of genome specific DNA markers of Brassica and evolution of amphidiploids and n=9 diploid species. Genome 33:131–142

Hu Q, Li Y, Mei D (2009) Introgression of genes from wild crucifers. In: Gupta SK (ed) Biology and breeding of crucifers. CRC Press, Boca Raton, FL, pp 261–283

Huang J, Rozwadowski K, Bhinu VS, Schäfer U, Hannoufa A (2008) Manipulation of sinapine, choline and betaine accumulation in Arabidopsis seed: Towards improving the nutritional value of the meal and enhancing the seedling performance under environmental stresses in oilseed crops. Plant Physiol Biochem 46:647–654. https://doi.org/10.1016/j.plaphy.2008.04.014

Huang H, Cui T, Zhang L et al (2020) Modifications of fatty acid profile through targeted mutation at BnaFAD2 gene with CRISPR/Cas9-mediated gene editing in *Brassica napus*. Theor Appl Genet 133:2401–2411. https://doi.org/10.1007/s00122-020-03607-y

Jagannath A, Sodhi YS, Gupta V et al (2011) Eliminating expression of erucic acid-encoding loci allows the identification of "hidden" QTL contributing to oil quality fractions and oil content in *Brassica juncea* (Indian mustard). Theor Appl Genet 122:1091–1103. https://doi.org/10.1007/s00122-010-1515-2

Jahangir M, Kim HK, Choi YH, Verpoorte R (2009) Health-affecting compounds in Brassicaceae. Compr Rev Food Sci Food Saf 8:31–43

Jain A, Bhatia S, Banga SS, Prakash S, Laxmikumaran M (1994) Potential use of random amplified polymorphic DNA (RAPD) technique to study genetic diversity in Indian mustard (*B. juncea*) and its relatedness to heterosis. Theor Appl Genet 88:116–122

James DW, Dooner HK (1991) Novel seed lipid phenotypes in combinations of mutants altered in fatty acid biosynthesis in Arabidopsis. Theor Appl Genet 82(4):409–412

James DW Jr, Lim E, Keller J, Plooy I, Ralston E et al (1995) Directed tagging of the arabidopsis *FATTY ACID ELONGATION1* (*FAE1*) gene with the maize transposon activator. Plant Cell 7: 309–319. https://doi.org/10.2307/3869853

Jha JK, Sinha S, Maiti MK, Basu A, Mukhopadhyay UK et al (2007) Functional expression of an acyl carrier protein (ACP) from *Azospirillum brasilense* alters fatty acid profiles in *Escherichia coli* and *Brassica juncea*. Plant Physiol Biochem 45:490–500. https://doi.org/10.1016/j.plaphy.2007.03.001

Jie M, Cheung WM, Yu V, Zhou Y, Tong PH et al (2014) Anti-proliferative activities of sinigrin on carcinogen-induced hepatotoxicity in rats. PLoS One 9:e110145. https://doi.org/10.1371/journal.pone.0110145

Kah M, Tufenkji N, White JC (2019) Nano-enabled strategies to enhance crop nutrition and protection. Nat Nanotechnol 14:532–540. https://doi.org/10.1038/s41565-019-0439-5

Kajla S, Mukhopadhyay A, Pradhan AK (2017) Development of transgenic *Brassica juncea* lines for reduced seed sinapine content by perturbing phenylpropanoid pathway genes. PLoS One 12. https://doi.org/10.1371/JOURNAL.PONE.0182747

Kang L, Qian L, Zheng M, Chen L, Chen H et al (2021) Genomic insights into the origin, domestication and diversification of *Brassica juncea*. Nat Genet 53:1392–1402. https://doi.org/10.1038/s41588-021-00922-y

Kanrar S, Venkateswari J, Dureja P, Kirti PB, Chopra VL (2006) Modification of erucic acid content in Indian mustard (*Brassica juncea*) by up-regulation and down-regulation of the *Brassica juncea Fatty Acid Elongation1* (*BjFAE1*) gene. Plant Cell Rep 25:148–155. https://doi.org/10.1007/s00299-005-0068-3

Kaplan NM (1989) The deadly quartet. Upper-body obesity, glucose intolerance, hypertriglyceridemia, and hypertension. Arch Intern Med 149:1514–1520. https://doi.org/10.1001/archinte.149.7.1514

Kapoor LD (1990) CRC handbook of ayurvedic medicinal plants. CRC Press, Boca Raton, FL, p 84

Katche E, Quezada-Martinez D, Katche EI (2019) Interspecific hybridization for Brassica crop improvement. Crop Breed Genet Genom:1–32. https://doi.org/10.20900/cbgg20190007

Katiyar RK (1997) Pusa Jai Kisan: a new high-yielding mustard variety developed through tissue culture. Indian Farming 45:6–7

Kaur G, Banga SK, Banga SS (2004) Introgression of desaturase suppressor gene(s) from *Brassica napus* L.to enhance oleic acid content in *Brassica juncea* L. Coss. In: Proceedings of the 4th international crop science congress. Brisbane, Australia

Khan S, Anwar S, Kuai J, Noman A, Shahid M, Din M, Ali A, Zhou G (2018) Alteration in yield and oil quality traits of winter rapeseed by lodging at different planting density and nitrogen rates. Sci Rep 8(1):634

Khare CP (2004) Indian herbal remedies: rational Western therapy, ayurvedic, and other traditional usage, botany. Springer, Berlin

Khattak AN, Wang T, Yu K, Yang R, Wan W et al (2019) Exploring the basis of 2-propenyl and 3-butenyl glucosinolate synthesis by QTL mapping and RNA-sequencing in *Brassica juncea*. PLoS One 14:e0220597. https://doi.org/10.1371/journal.pone.0220597

Khulbe RK, Pant DP, Saxena N (2000) Variability, heritability and genetic advance in Indian mustard [*Brassica juncea* (L.) Czern & Coss]. Crop Res (Hisar) 20:551–552

Kim MK, Park JHY (2009) Cruciferous vegetable intake and the risk of human cancer: epidemiological evidence. Proc Nutr Soc 68:103–110

King A (2017) Technology: the future of agriculture. Nature 544:S21–S23

Kirk JTO, Oram RN (1978) Mustards as possible oil and protein crops for Australia (Indian mustard, *Brassica juncea*, white mustard, *Sinapsis alba*). J Austral Inst Agric Sci 44: 143–156

Kirti PB, Narasimhulu SB, Prakash S, Chopra VL (1992) Somatic hybridization between *Brassica juncea* and *Moricandia arvensis* by protoplast fusion. Plant Cell Rep 11(5–6):318–321

Kondra ZP, Thomas PM (1975) Inheritance of oleic, linoleic and linolenic acids in seed oil of rapeseed (*Brassica napus*). Can J Plant Sci 55:205–210

Krishna P, Ghose SK (1992) Heterosis in relation to genetic divergence in rapessed and mustard. J Oilseeds Res 9:169–174

Krzymanski J, Downey RK (1969) Inheritance of fatty acid composition in winter forms of rapeseed, *Brassica napus*. Can J Plant Sci 49:313–319. https://doi.org/10.4141/cjps69-053

Kumar S, Mawlong I, Rani R (2020) Biofortification of Brassicas for quality improvement. In: Wani SH, Thakur AK, Khan YJ (eds) Brassica improvement. Springer, Cham, pp 127–145. https://doi.org/10.1007/978-3-030-34694-2_7

Kumar D, Rajwanshi R, Singh P et al (2022) Pyramiding of γ-TMT and gly I transgenes in *Brassica juncea* enhances salinity and drought stress tolerance. Physiol Plant 174:e13618. https://doi.org/10.1111/ppl.13618

Labana KS, Banga SS (1984) Floral biology in Indian mustard (*Brassica juncea* L.). Genetica Agraria 38:131–138

Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V et al (2013) Acyl-lipid metabolism. In: The arabidopsis book/American Society of Plant Biologists, p 11

Lietzow J (2021) Biologically active compounds in mustard seeds: a toxicological perspective. Foods 10(9):2089

Lionneton E, Ravera S, Sanchez L et al (2002) Development of an AFLP-based linkage map and localization of QTLs for seed fatty acid content in condiment mustard (*Brassica juncea*). Genome 45:1203–1215. https://doi.org/10.1139/g02-095

Lionneton E, Aubert G, Ochatt S, Merah O (2004) Genetic analysis of agronomic and quality traits in mustard (Brassica juncea). Theor Appl Genet 109:792–799

Liu S, Liu Y, Yang X, Tong C, Edwards D et al (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun 5:1–11. https://doi.org/10.1038/ncomms4930

Love HR, Rakow G, Raney JP, Downey RK (1990) Genetic control of 2-propenyl and 3-butenyl glucosinolate synthesis in mustard. Can J Plant Sci 70:425–429

Luciano FB, Holley RA (2009) Enzymatic inhibition by allyl isothiocyanate and factors affecting its antimicrobial action against Escherichia coli O157:H7. Int J Food Microbiol 131:240–245. https://doi.org/10.1016/j.ijfoodmicro.2009.03.005

Lüthy J, Carden B, Friederich U, Bachmann M (1984) Goitrin – a nitrosatable constituent of plant foodstuffs. Experientia 40:452–453. https://doi.org/10.1007/BF01952381

Lysak MA, Koch MA, Pecinka A, Schubert I (2005) Chromosome triplication found across the tribe Brassiceae. Genome Res 15:516–525. https://doi.org/10.1101/gr.3531105

Mahmood T, Ekuere U, Yeh F, Good AG, Stringam GR (2003a) Molecular mapping of seed aliphatic glucosinolates in *Brassica juncea*. Genome 46(5):753–760. https://doi.org/10.1139/g03-051

Mahmood T, Ekuere U, Yeh F et al (2003b) RFLP linkage analysis and mapping genes controlling the fatty acid profile of *Brassica juncea* using reciprocal DH populations. Theor Appl Genet 107:283–290. https://doi.org/10.1007/s00122-003-1244-x

Marwede V, Gul MK, Becker HC, Ecke W (2005) Mapping of QTL controlling tocopherol content in winter oilseed rape. Plant Breed 124:20–26

Mason AS, Higgins EE, Snowdon RJ et al (2017) A user guide to the Brassica 60K Illumina InfiniumTM SNP genotyping array. Theor Appl Genet 130:621–633. https://doi.org/10.1007/s00122-016-2849-1

Mathur S, Paritosh K, Tandon R, Pental D, Pradhan A (2022) Comparative analysis of seed transcriptome and coexpression analysis reveal candidate genes for enhancing seed size/weight in *Brassica juncea*. Front Genet 13. https://doi.org/10.3389/fgene.2022.814486

Mawlong I, Kumar MS, Kandpal BK, Premi OP, Gurung B, Singh D (2017) Meal and oil quality among genotypes of Indian mustard (*Brassica juncea*) varies under recommended dose of nitrogen fertilizer. Appl Ecol Environ Res 15:1427–1445

Mawlong I, Kumar S, Kandpal BK, Premi OP, Joshi A, Gurung B (2018) Comparing oil color and oxidative stability among mustard genotypes under nitrogen fertilization. Agron J 110:1–11

McNaughton SA, Marks GC (2003) Development of a food composition database for the estimation of dietary intakes of glucosinolates, the biologically active constituents of cruciferous vegetables. Br J Nutr 90:687–697

Meena CP, Chauhan JS, Singh M, Meena ML, Singh KH (2014) Evaluation of Indigenous and Exotic Germplasm of Indian mustard [*Brassica juncea* (L.) Czern & Coss] for morpho physiological and quality characters. Indian J Plant Genet Resour 27:118–124

Morales-Lopez J, Centeno-Alvarez M, Nieto-Camacho A, Lopez MG, Perez-Hernandez E et al (2017) Evaluation of antioxidant and hepatoprotective effects of white cabbage essential oil. Pharm Biol 55:233–241

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res 35. https://doi.org/10.1093/nar/gkm321

Mueller LA, Zhang P, Rhee SY (2003) AraCyc: a biochemical pathway database for Arabidopsis. Plant Physiol 132:453–460. https://doi.org/10.1104/pp.102.017236

Müller J, Sonntag K, Rudloff E (2001) Somatic hybridization between *Brassica* spp. and *Raphanus sativus*. Acta Hortic 560:219–220. https://doi.org/10.17660/ActaHortic.2001.560.41

Mun JH, Kwon SJ, Yang TJ, Seol YJ, Jin M et al (2009) Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. Genome Biol 10. https://doi.org/10.1186/gb-2009-10-10-r111

Nagamine A, Hiroshi E (2022) Genome editing for improving crop nutrition. Front Gen Edi 4: 850104

Nelson MN, Lilley JM, Helliwell C, Taylor CM, Siddique KH et al (2016) Can genomics assist the phenological adaptation of canola to new and changing environments? Crop Pasture Sci 67(4): 284–297

Nour-Eldin HH, Madsen SR, Engelen S, Jørgensen ME, Olsen CE et al (2017) Reduction of antinutritional glucosinolates in Brassica oilseeds by mutation of genes encoding transporters. Nat Biotechnol 35:377–382. https://doi.org/10.1038/nbt.3823

Ogura H (1968) Studies on the new male sterility in Japanes radish with special reference to the utilization of this sterility towards the practical raising of hybrid seed. Mem Fac Agric Kagoshima Univ 6:39–78

Okulicz M (2010) Multidirectional time-dependent effect of Sinigrin and allyl Isothiocyanate on metabolic parameters in rats. Plant Foods Hum Nutr 65:217–224. https://doi.org/10.1007/s11130-010-0183-3

Padmaja LK, Agarwal P, Gupta V, Mukhopadhyay A, Sodhi YS et al (2014) Natural mutations in two homoeologous *TT8* genes control yellow seed coat trait in allotetraploid *Brassica juncea* (AABB). Theor Appl Genet 127:339–347. https://doi.org/10.1007/s00122-013-2222-6

Pandey S, Kabdal M, Tripathi MK (2013) Study of inheritance of erucic acid in Indian mustard (*Brassica juncea* L. Czern & Coss). Octa J Biosci 1(1)

Panjabi P, Jagannath A, Bisht NC, Padmaja KL, Sharma S et al (2008) Comparative mapping of *Brassica juncea* and *Arabidopsis thaliana* using Intron Polymorphism (IP) markers: homoeologous relationships, diversification and evolution of the A, B and C Brassicagenomes. BMC Genomics 9:113

Panjabi P, Yadava SK, Kumar N, Bangkim R, Ramchiary N (2019) Breeding *Brassica juncea* and *B. rapa* for sustainable oilseed production in the changing climate: progress and prospects. In: Kole C (eds) Genomic designing of climate-smart oilseed crops. Springer, Cham. https://doi.org/10.1007/978-3-319-93536-2_6

Panjabi-Massand P, Yadava SK, Sharma P, Kaur A, Kumar A et al (2010) Molecular mapping reveals two independent loci conferring resistance to *Albugo candida* in the East European germplasm of oilseed mustard *Brassica juncea*. Theor Appl Genet 121:137–145. https://doi.org/10.1007/s00122-010-1297-6

Pant SC, Singh P (2001) Genetic variability in Indian mustard. Agric Sci Dig 21:28–30

Pant P, Hamsa S, Kaur J (2022) Advances in breeding strategies for improving stress tolerance in Brassicas. In: The *Brassica juncea* genome. Springer, pp 439–469

Paritosh K, Yadava SK, Singh P, Bhayana L, Mukhopadhyay A et al (2021) A chromosome-scale assembly of allotetraploid *Brassica juncea* (AABB) elucidates comparative architecture of the A and B genomes. Plant Biotechnol J 19:602–614. https://doi.org/10.1111/pbi.13492

Parkin IAP, Gulden SM, Sharpe AG, Lukens L, Trick M et al (2005) Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. Genetics 171:765–781. https://doi.org/10.1534/genetics.105.042093

Pental D, Pradhan AK, Sodhi YS, Mukhopadhyay A (1993) Variation amongst *Brassica juncea* cultivars for regeneration from hypocotyl explants and optimization of conditions for Agrobacterium-mediated genetic transformation. Plant Cell Rep 12:462–467. https://doi.org/10.1007/BF00234713

Potts DA, Males DR (1999) Inheritance of fatty acid composition in *Brassica juncea*. In: 10th International Rapeseed Congress, Canberra, Australia, p 9

Pradhan AK, Pental D (2011) Genetics of *Brassica juncea* L. In: Schmidt R, Bancroft I (eds) Genetics and genomics of the Brassicaceae, vol 9. Springer, New York, pp 323–346

Pradhan AK, Sodhi YS, Mukhopadhyay A, Pental D (1993) Heterosis breeding in Indian mustard (*Brassica juncea* L. Czern & Coss): Analysis of component characters contributing to heterosis for yield. Euphytica 69:219–229. https://doi.org/10.1007/BF00022368

Pradhan A, Gupta V, Mukhopadhyay A et al (2003) A high-density linkage map in *Brassica juncea* (Indian mustard) using AFLP and RFLP markers. Theor Appl Genet 106:607–614. https://doi.org/10.1007/s00122-002-1083-1

Prakash S, Hinata K (1980) Taxonomy, cytogenetics and origin of crop brassicas – a review. Oper Bot 55:1–57

Prakash S, Bhat SR, Fu T (2009) Wild germplasm and male sterility. In: Biology and breeding of crucifers, pp 113–127

Pushpa HD, Yadava DK, Singh N et al (2016) Validation of molecular markers linked to low glucosinolate QTLs for marker assisted selection in Indian mustard (*Brassica juncea* L. Czern & Coss). Indian J Genet Plant Breed 76:64. https://doi.org/10.5958/0975-6906.2016.00009.2

Rabbani MA, Iwabuchi A, Murakami Y, Suzuki T, Takayanagi K (1998) Phenotypic variation and the relationships among mustard (*Brassica juncea* L.) germplasm from Pakistan. Euphytica 101(3):357–366

Rakow G, Woods DL (1987) Outcrossing in rape and mustard under Saskatchewan prairie conditions. Can J Plant Sci 67:147–151

Ramchiary N, Bisht NC, Gupta V, Mukhopadhyay A, Arumugam N, Sodhi YS, Pental D, Pradhan AK (2007) QTL analysis reveals context-dependent loci for seed glucosinolate trait in the oilseed *Brassica juncea*: importance of recurrent selection backcross scheme for the identification of 'true' QTL Theoretical and Applied Genetics 116(1):77–85. https://doi.org/10.1007/s00122-007-0648-4

Rastogi T, Reddy KS, Vaz M, Spiegelman D, Prabhakaran et al (2004) Diet and risk of ischemic heart disease in India. Am J Clin Nutr 79:582e592

Ravi MS, Geethanjali, Sameeyafarheen F, Maheswaran M (2003) Molecular marker based genetic diversity analysis in rice (Oryza sativa L.) using RAPD and SSR markers. Euphytica 133(2): 243–252

Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D et al (2003) The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. Nucleic Acids Res 31:224–228. https://doi.org/10.1093/NAR/GKG076

Ripley VL, Thompson SA, Ehlert ZC (2014) Omega-9 quality *Brassica juncea*. US Patent 8,637,740, 28 Jan 2014

Röbbelen G (1990) Mutation breeding for quality improvement a case study for oilseed crops. Plant Mutat Breed Crop Improv 2:3–30

Röbbelen G, Nitsch A (1975) Genetical and physiological investigations on mutants for polyenoic fatty acids in rapeseed, *Brassica napus* L. I. Selection and description of new mutants. Plant Breed 75:93–105

Rohilla P, Naik R, Acharya R (2017) A look into the contribution of raja nighantu, an ayurvedic lexicon of 14th century AD, towards drug safety. Glob J Res Med Plants Indig Med 6:34e53

Rosa EAS, Heaney RK, Fenwick GR, Portas CAM (1997) Glucosinolates in crop plants. Hortic Rev 19:199–215

Rout K, Sharma M, Gupta V et al (2015) Deciphering allelic variations for seed glucosinolate traits in oilseed mustard (*Brassica juncea*) using two bi-parental mapping populations. Theor Appl Genet 128:657–666. https://doi.org/10.1007/s00122-015-2461-9

Rout K, Yadav BG, Yadava SK, Mukhopadhyay A, Gupta V et al (2018) QTL landscape for oil content in *Brassica juncea*: analysis in multiple bi-parental populations in high and "0" erucic background. Front Plant Sci 871:1448. https://doi.org/10.3389/fpls.2018.01448

Sanlier N, Guler Saban M (2018) The benefits of *Brassica* vegetables on human health. J Human Health Res 1:104

Savadi S, Naresh V, Kumar V, Bhat SR (2015) Seed-specific overexpression of Arabidopsis *DGAT1* in Indian mustard (*Brassica juncea*) increases seed oil content and seed weight. Botany 94:177–184. https://doi.org/10.1139/cjb-2015-0218

Schranz ME, Lysak MA, Mitchell-Olds T (2006) The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. Trends Plant Sci 11:535–542. https://doi.org/10.1016/j.tplants.2006.09.002

Sharafi Y, Majidi MM, Goli SAH, Rashidi F (2015) Oil content and fatty acids composition in Brassica species. Int J Food Prop 18(10):2145–2154

Sharma R, Prajapati PK (2016) Antidiabetic leads from ayurvedic medicinal plants. Int J Adv Complement Tradit Med 2:24e41

Sharma A, Mohapatra T, Sharma RP (1994) Molecular mapping and character tagging in *brassica juncea* – i. degree, nature and linkage relationship of RFLPs and their association with quantitative traits. J Plant Biochem Biotechnol 3:85–89. https://doi.org/10.1007/BF03321956

Sharma R, Aggarwal RA, Kumar R et al (2002) Construction of an RAPD linkage map and localization of QTLs for oleic acid level using recombinant inbreds in mustard (*Brassica juncea*). Genome 45:467–472. https://doi.org/10.1139/g02-001

Sharma M, Mukhopadhyay A, Gupta V, Pental D, Pradhan AK (2016) *BjuB.CYP79F1* regulates synthesis of propyl fraction of aliphatic glucosinolates in oilseed mustard *Brassica juncea*: functional validation through genetic and transgenic approaches. PLoS One 11:1–17. https://doi.org/10.1371/journal.pone.0150060

Sharma HK, Singh VV, Kumar A, Meena HS, Sharma P, Rai PK (2022) Genepools of Brassica. In: Kole C, Mohapatra T (eds) The *Brassica juncea* Genome. Compendium of plant genomes. Springer, Cham. https://doi.org/10.1007/978-3-030-91507-0_4

Shen S, Tang Y, Zhang C, Yin N, Mao Y et al (2021) Metabolite profiling and transcriptome analysis provide insight into seed coat color in *Brassica juncea*. Int J Mol Sci 22:7215. https://doi.org/10.3390/ijms22137215

Shin JM, Eunbi L, Yoon SC, Chu WN (2021) Cancer-preventive effect of phenethyl isothiocyanate through tumor microenvironment regulation in a colorectal cancer stem cell xenograft model. Phytomedicine 84:153493

Shyam C, Tripathi MK, Tiwari S et al (2021) In vitro production of Somaclones with decreased erucic acid content in Indian mustard [*Brassica juncea* (L.) Czern & Coss]. Plants 10:1297. https://doi.org/10.3390/plants10071297

Singh BP, Chowdhury RK (1983) Correlation and path coefficient analysis of seed yield and oil content in mustard (*Brassica juncea*). Can J Genet Cytol 25:312–317

Singh BD, Singh AK (2015) Marker-assisted plant breeding: principles and practices. Springer India, New Delhi

Singh M, Singh N, Bhandari DK (1980) Interaction of selenium and sulfur on the growth and chemical composition of Raya. Soil Sci 129(4):238–244

Singh B, Singh J, Kumar A, Yadav YP, Singh S (1999) Response of Brassicas to sulphur for yield. Indian J Agric Sci 69(6):427–431

Singh S, Mohapatra T, Singh R et al (2013) Mapping of QTLs for oil content and fatty acid composition in Indian mustard [*Brassica juncea* (L.) Czern. and Coss.]. J Plant Biochem Biotechnol 22:80–89. https://doi.org/10.1007/s13562-012-0113-6

Singh VK, Avtar R, Kumari N, Kumar R (2021) Assessment of genetic diversity and population structure in Indian mustard (*Brassica juncea* L.) using SSR markers. J Environ Biol 42:396–405

Sinha S, Jha JK, Maiti MK, Basu A, Mukhopadhyay UK et al (2007) Metabolic engineering of fatty acid biosynthesis in Indian mustard (*Brassica juncea*) improves nutritional quality of seed oil. Plant Biotechnol Rep 1:185–197. https://doi.org/10.1007/s11816-007-0032-5

Sivaraman I, Arumugam N, Sodhi YS, Gupta V, Mukhopadhyay A et al (2004) Development of high oleic and low linoleic acid transgenics in a zero erucic acid *Brassica juncea* L. (Indian mustard) line by antisense suppression of the *fad2* gene. Mol Breed 13:365–375. https://doi.org/10.1023/B:MOLB.0000034092.47934.d6

Small E (2006) Culinary herbs. National Research Council Canada, Ottawa, p 221e228

Smith JSC, Smith OS (1989) The description and assessment of distances between inbred lines of maize: the utility of morphological, biochemical and genetic descriptors and a scheme for the testing of distinctiveness between inbred lines. Maydica 34:151–161

Sodhi YS, Mukhopadhyay A, Arumugam N, Verma JK, Gupta V, Pental D (2002) Genetic analysis of total glucosinolate in crosses involving a high glucosinolate Indian variety and low glucosinolate line of *Brassica juncea*. Plant Breed 121:508–511

Sodhi YS, Chauhan A, Verma JK, Arumugam N, Mukhopadhyay A, Gupta V, Pental D, Pradhan AK (2006) A new cytoplasmic male sterility system for hybrid seed production in Indian oilseed mustard (*Brassica juncea*). Theor Appl Genet 114:93–99

Song KM, Osborn TC, Williams PH (1990) Brassica taxonomy based on nuclear restriction fragment length polymorphism (RFLPs). 3. Genome relationship in Brassica and related genera and the origin of *B. oleracea* and *B.rapa* (syn. *campestris*). Theor Appl Genet 79:497–506

Srivastava A, Gupta V, Pental D, Pradhan AK (2001) AFLP-based genetic diversity assessment amongst agronomically important natural and some newly synthesized lines of *Brassica juncea*. Theor Appl Genet 102:193–199. https://doi.org/10.1007/s001220051635

Srivastava RK, Singh RB, Pujarula VL et al (2020) Genome-wide association studies and genomic selection in pearl millet: advances and prospects. Front Genet 10. https://doi.org/10.3389/fgene.2019.01389

Stoutjesdijk PA, Hurlestone C, Singh SP, Green AG (2000) High-oleic acid Australian *Brassica napus* and *B. juncea* varieties produced by co-suppression of endogenous *Δ12-desaturases*. Biochem Soc Trans 28:938–940. https://doi.org/10.1042/BST0280938

Stringam GR, Thiagarajah MR (1995) Inheritance of alkenyl glucosinolates in traditional and microsporederived doubled haploid populations of *Brassica juncea* L. Czern and Coss. In: Proceedings of 9th international rapeseed congress, 4–7 July 1995. Cambridge, UK, pp 804–806

Sun FJ (1943) Hybrid vigor in Brassica. J Agric Assoc China 175:35–58

Tandayu E, Borpatragohain P, Mauleon R, Kretzschmar T (2022) Genome-wide association reveals trait loci for seed Glucosinolate accumulation in Indian mustard (*Brassica juncea* L.). Plants 11:364. https://doi.org/10.3390/plants11030364

The Arabidopis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815. https://doi.org/10.1038/35048692

Thomas J, Kuruvilla KM, Hrideek TK (2012) Mustard. In: Handbook of herbs and spices. Woodhead Publishing, pp 388–398

Thurling N, Depittayanan V (1992) EMS Induction of early flowering mutants in spring rape (*Brassica napus*). Plant Breed 108:177–184

Tiwari AS (1995) Improved quality of oil and meal in Indian mustard. In: Rapeseed today and tomorrow, 9th international rapeseed congress. Cambridge, pp 434–436

U.S. Department of Agriculture, Agricultural Research Service. FoodData Central, 2019

UN (1935) Genome analysis in Brassica with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Jpn J Bot 7:389–452

van Doorn HE, van der Kruk GC, van Holst GJ, Raaijmakers-Ruijs NC, Postma E et al (1998) The glucosinolates sinigrin and progoitrin are important determinants for taste preference and bitterness of Brussels sprouts. J Sci Food Agric 78(1):30–38

Varshney RK, Roorkiwal M, Sorrells ME (2017) Genomic selection for crop improvement: an introduction. In: Varshney R, Roorkiwal M, Sorrells M (eds) Genomic selection for crop improvement. Springer, Cham, pp 1–6

Vaughan G (1977) A multidisciplinary study of the taxonomy and origin of Brassica crops. Bioscience 27:35–40

Vaughan JG, Hemingway JS, Schofield HJ (1963) Contributions to a study of variation in *Brassica juncea* Coss. & Czern. J Linn Soc Bot 58:435–447

Vavilov NI (1949) The origin, variation, immunity and breeding of cultivated plants. Chron Bot 13:1–364

Venkateswari J, Kanrar S, Kirti PB et al (1999) Molecular Cloning and Characterization of FATTY ACID ELONGATION1 (*BjFAE1*) Gene of *Brassica juncea*. J Plant Biochem Biotechnol 8:53–55. https://doi.org/10.1007/BF03263058

Verma K, Tripathi MK, Tiwari S, Tripathi N (2021) Analysis of genetic diversity among *Brassica juncea* genotypes using morphophysiological and SSR markers. Int J Curr Microbiol App Sci 10(01):1108–1117

Wang X, Wang H, Wang J, Sun R, Wu J et al (2011) The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet 43:1035–1040. https://doi.org/10.1038/ng.919

Wang T, Liang H, Yuan Q (2012) Separation of sinigrin from Indian mustard (*Brassica juncea* L.) seed using macroporous ion-exchange resin. Korean J Chem Eng 29:396–403. https://doi.org/10.1007/s11814-011-0175-5

Wang G, Zhang X, Huang W, Xu P, Lv Z et al (2021) Increased seed number per silique in *Brassica juncea* by deleting cis-regulatory region affecting *BjCLV1* expression in carpel margin meristem. Plant Biotechnol J 19:2333–2348. https://doi.org/10.1111/pbi.13664

White PJ, Brown PH (2010) Plant nutrition for sustainable development and global health. Ann Bot 105(7):1073–1080

Woods DL, Potts DA, Males DR (1999) Genetic control of C18 fatty acid in *Brassica juncea*. In: 10th International Rapeseed Congress, Canberra, Australia, p 142

Wu XM, Chen BY, Lu G et al (2009) Genetic diversity in oil and vegetable mustard (*Brassica juncea*) landraces revealed by SRAP markers. Genet Resour Crop Evol 56:1011. (2009). https://doi.org/10.1007/s10722-009-9420-8

Yadav RC, Kamada H, Kikuchi F (1991) Genotypic and media effects on plant regeneration from cotyledonary explants of *Brassica juncea* (L.) Coss & Czern. Ann Biol 7:119–124

Yadav R, Singh R, Kumar S, Prasad TV, Bharadwaj R (2017) Genetic diversity among indigenous germplasm of *Brassica juncea* (L.) Czern and Coss, using agro-morphological and phenological traits. Proc Natl Acad Sci India Sect B Biol Sci 87(4):1125–1131

Yadava SK, Arumugam N, Mukhopadhyay A, Sodhi YS, Gupta V et al (2012) QTL mapping of yield-associated traits in *Brassica juncea*: Meta-analysis and epistatic interactions using two different crosses between east European and Indian gene pool lines. Theor Appl Genet 125:1553–1564. https://doi.org/10.1007/s00122-012-1934-3

Yan W, Zhao H, Yu K et al (2020) Development of a multiparent advanced generation intercross (MAGIC) population for genetic exploitation of complex traits in *Brassica juncea*: Glucosinolate content as an example. Plant Breed 139:779–789. https://doi.org/10.1111/pbr.12820

Yang J, Song N, Zhao X et al (2014) Genome survey sequencing provides clues into glucosinolate biosynthesis and flowering pathway evolution in allotetrapolyploid *Brassica juncea*. BMC Genomics 15:107. https://doi.org/10.1186/1471-2164-15-107

Yang J, Liu D, Wang X, Ji C, Cheng F et al (2016) The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. Nat Genet 48:1225–1232. https://doi.org/10.1038/ng.3657

Yang J, Wang J, Li Z et al (2021) Genomic signatures of vegetable and oilseed allopolyploid *Brassica juncea* and genetic loci controlling the accumulation of glucosinolates. Plant Biotechnol J 19:2619–2628. https://doi.org/10.1111/pbi.13687

Yao K, Bacchetto RG, Lockhart KM, Friesen LJ, Potts DA et al (2003) Expression of the Arabidopsis *ADS1* gene in *Brassica juncea* results in a decreased level of total saturated fatty acids. Plant Biotechnol J 1:221–229. https://doi.org/10.1046/j.1467-7652.2003.00021.x

Yu JC, Jiang ZT, Li R, Chan SM (2003) Chemical composition of the essential oils of *Brassica juncea* (L) CossGrown in different regions, Hebei, Shaanxi and Shandong, of China. J Food Drug Anal 11(1)

Yusuf MA, Sarin NB (2007) Antioxidant value addition in human diets: genetic transformation of *Brassica juncea* with γ-TMT gene for increased α-tocopherol content. Transgenic Res 16:109–113. https://doi.org/10.1007/s11248-006

Zhang Y (2010) Allyl isothiocyanate as a cancer chemopreventive phytochemical. Mol Nutr Food Res 54:127–135

Zhang M, Viennois E, Xu C, Merlin D (2016) Plant derived edible nanoparticles as a new therapeutic approach against diseases. Tissue Barriers 4:e1134415. https://doi.org/10.1080/21688370.2015.1134415

Nutraceutomics of the Ancient Oilseed Crop Sesame (*Sesamum indicum* L.)

Yinghui Duan, Hongmei Miao, Ming Ju, Chun Li, Hengchun Cao, and Haiyang Zhang

Contents

Abstract

Sesame is an ancient oilseed crop and is widely cultivated in the tropical and subtropical regions of the world. It is well known for its high-quality nutritional seeds with abundant fatty acids (~55%), proteins (~20%), various vitamins and

Y. Duan · H. Miao · M. Ju · C. Li · H. Cao · H. Zhang (✉)
Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, Henan, China

Henan Key Laboratory of Specific Oilseed Crops Genomics (Henan Sesame Research Center, Henan Academy of Agricultural Sciences), Zhengzhou, Henan, China

Henan International Joint Laboratory of Specific Oilseed Crops Improvement, Zhengzhou, Henan, China
e-mail: zhanghaiyang@zzu.edu.cn

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_17

minerals, and natural antioxidants, such as sesamin, sesamolin, and tocopherols. Sesame is consumed directly as the edible seed and oil and applied for nutrition and medicine industry as the functional food and nutraceuticals because of its anti-oxidative, anti-inflammatory, hypolipidemic, cardioprotective, neuroprotective, and anticarcinogenic effects. In the past four decades, a great deal of new sesame varieties with high-yield potential and elite agronomic traits including the high resistance to biotic and abiotic stresses and the high content of seed nutrients were bred through high-efficient breeding techniques and released for sesame industry. Achievement of the Sesame Genome Project with huge amount of sesame genomic data impedes the rapid development of genomics and molecular genetics research in sesame. As a result, a list of invaluable genetic resources, including molecular makers, genetic maps, quantitative trait loci (QTLs), and functional genes, were detected to decipher the genetic basis of the especially nutrition-related traits and improve the molecular breeding strategies for sesame. We herein summarize the available information about sesame nutraceutomics and molecular breeding research. We systematically introduced the technological progresses mainly in new germplasm creation, highly efficient marker-assisted breeding, and genetic transformation for sesame and forecasted the development of sesame nutrition mechanism studies so as to meet the consumer's demands.

Keywords

Sesame · Nutrition · Bioactive compound · Genomics · Molecular breeding

1 Introduction

Sesame (*Sesamum indicum* L., 2n = 26) belongs to the *Sesamum* genus of the Pedaliaceae family and is the sole cultivated species in *Sesamum*. Sesame is an ancient oilseed crop with high oil quality. Its domestication and cultivation history could be traced back to the Harappa and Anatolian eras on the Indian subcontinent 5000 years ago (Bedigian 2003). At present, sesame is widely planted in the tropical and subtropical regions of Africa, Asia, and Southern America (Pathak et al. 2017). In 2020, the total harvest area of the world sesame reached 13.97 million hectares. The total product was 6.8 million tons, while the average yield was low at 487.2 kg per hectare for the low production technology in the main production countries (Table 1) (FAOSTAT, 2020). During the past decade, the total production of the world sesame was consistently rising mainly due to the contribution of harvest area increase especially in Africa (FAOSTAT, 2011–2020).

Sesame is a traditional and manually produced crop, but meets the continuous global demand for high-quality vegetable oils during the past decades. On the other hand, sesame plays an important role as a cash crop in the development of local economy and cultivation structure (Zhang et al. 2021a) as the major production countries, including Sudan, Myanmar, Tanzania, India, Nigeria, and China are developing countries (Table 1). In 2020, 7 of the top 10 sesame production countries

Table 1 Sesame production statistics of the world and the top 10 production countries in 2020

Country	Area harvested (ha)	Production (tons)	Yield (kg/ha)
World	13,965,844	6,803,824	487.2
Sudan	5,173,521	1,525,104	294.8
Myanmar	1,500,000	740,000	493.3
Tanzania	960,000	710,000	739.6
India	1,520,000	658,000	432.9
Nigeria	621,413	490,000	788.5
China	278,652	449,452	1613.0
Burkina Faso	450,000	270,000	600.0
Ethiopia	369,897	260,258	703.6
Chad	392,241	202,074	515.2
South Sudan	608,159	189,721	312.0
Others	2,091,961	1,309,215	625.8

The data are cited from FAO dataset (www.fao.org/statistics/en/)

were from Africa. Sudan was the biggest sesame producer, while China presented the highest yield level in the world.

Sesame is traditionally produced for edible seed production and oil crash. Sesame seeds contain high oil content (~55%), and the monounsaturated and polyunsaturated fatty acid compose more than 80% of the oil. Besides plenty of fatty acids, abundant proteins and dietary fibers, various vitamins and minerals and natural antioxidants, such as sesamin, sesamol, sesamolin, and tocopherol derivatives improve the quality of sesame seeds and oil (Anilakumar et al. 2010; Pathak et al. 2017). Sesame seeds have been regarded as "the queen of oilseed crop seeds" as possessing the abundance of nutrients. Many reports have demonstrated that the specific unsaturated fatty acids with lignans in sesame seeds have the health-promoting effects on preventing cardiovascular diseases, suppressing age-related neurodegenerative diseases, reducing the incidence of cancer, and improving wound healing, because of their antiatherogenic, antithrombotic, anti-inflammatory, hypolipidemic, cardioprotective, neuroprotective, antiaging effects, and antioxidative activity (Shenoy et al. 2011; Ray and Katyal 2016; Pathak et al. 2017; Andargie et al. 2021; Langyan et al. 2022). Thus, sesame is being used as the functional food and nutraceuticals on account of the specific health-related properties. In addition, main chemical composition of sesame is also applied as pharmaceutical substances, antifungal, insecticides antiseptics, disinfectants, and cosmetics, even as biofuel in pharmaceutic and industrial areas (Pathak et al. 2017; Zhang et al. 2021a).

Sesame originated in Africa as most of the wild relatives of cultivated sesame are found in the African continent (Nimmakayala et al. 2011; Zhang et al. 2021b). An Indian wild species *S. malabaricum* (2n = 26) has been regarded as the closest species to *S. indicum* and might be the direct progenitor of sesame (Bedigian 2003; Andargie et al. 2021; Zhang et al. 2021a). In the past decades, sesame is regarded as an orphan crop (Dossa et al. 2017). One reason is that the belonged small Pedaliaceae family of sesame is remote from other big families with oilseed crops,

such as Cruciferae (oilseed rape), Asteraceae (sunflower), and Fabaceaae family (soybean and peanut) in the plant phylogenetic relationship (Zhang et al. 2013). Few related species could be directly used as references for sesame genetics and breeding research. The second is that the genetic basis of sesame is relatively narrow as sesame is the sole cultivated species in the *Sesamum* genus. Increasing the yield potential with high resistance to biotic and abiotic stresses seems too difficult for sesame in the near future. At present, sesame contributes only about 1% of the global oilseed production and the competition capability with other major oilseed crops including rapeseed, soybean, peanut, and sunflower is still low. However, elite agronomic traits and the specific species characteristics determine the important academic position of sesame. Long cultivation history indicated the specific phylogenetic position of sesame in plants. The small genome size (~369 Mb) (Zhang et al. 2013) close to rice and relatively simple genome structure facilitates the genomics analysis of sesame and the wild *Sesamum* species. Moreover, high oil content and seed quality, rapid growth, and abundant seeds of a progeny also help to perform genetics and breeding research in sesame. Thus, sesame is regarded as a model of oilseed crop for its academic importance (Zhang et al. 2019).

In recent years, numerous studies have been implemented to improve agronomic traits and production through traditional and modern breeding methods in sesame. Especially, with the complete genome sequencing and the release of enormous genomic resources, sesame research has been led into the "Omics" era since 2013 (Dossa et al. 2017; Wang et al. 2014; Wei et al. 2015; Zhang et al. 2013; Zhang et al. 2021a). So much invaluable molecular resources, including omics data, genome information, molecular markers, and functional genes, are being applied to decipher the genetic basis of the seed yield and quality traits and to speed up sesame variety improvement. Thus, in this chapter, we give a systematical account of sesame nutritional compounds and their related genetic bases and regulation mechanisms. We summarize the main theoretical and technological achievements in sesame seed quality breeding. Some updated data and unpublished information obtained in the past a few years are shown in the chapter so as to facilitate readers understanding the key topic points.

2 Nutrient Components in Sesame

Sesame is the highest oil content in annual herbs. In sesame seed, the main nutrients include oil, protein, edible fibers, vitamins, antioxidants, and other components for providing energy and health-related nutrient (Prakash and Naik 2014). For the diverse sesame germplasm under different cultivation conditions, the oil and protein contents vary from 29.5–62.7% and 12.9–31%, respectively (Anilakumar et al. 2010; Zhang et al. 2021). The contents of lignans, mainly including sesamin and sesamolin, range from 0.517 to 15.832 mg/g, which can keep sesame oil stable and high quality (Zhang et al. 2021a). In addition, the carbohydrates also present in sesame seeds with the highest content of 17.4%. The digestible fiber content is high at 9.8% and exists mainly in sesame seed hull. Noticeably, the polysaccharide takes

up the total seed weight of 6.5%, which are mainly composed of galacturonic acid, glucuronic acid, rhamnose and glucose, and possesses the antioxidant activity (Zhang et al. 2022). Moreover, sesame seed contains abundant vitamins, such as pantothenic acid, vitamin E, folic acid, lactochrome (B2), niacin, thiamine (B1), and vitamin C. Sesame seeds and leaves also are excellent sources of minerals, such as magnesium, potassium, calcium, copper, iron, zinc, and selenium (Zhang et al. 2021a).

With the above beneficial nutritional composition, sesame seeds are widely used to produce raw or roasted sesame seeds, sesame oil, sesame paste, sesame flour, confectionery, and bakery ingredient for food production. In specific regions, sesame products also exhibit the local food culture. In China, sesame oil is even used as a substitute for olive oil. Sesame meal with the high protein content is consumed as an important food in India. Tahini, a paste made from roasted hulled sesame seeds, is used as a traditional sugar ingredient in Eastern Asia.

2.1 Sesame Oil Composition

Sesame oil is composed of more than 95% triacylglycerols and a small amount of diacylglycerols, free fatty acids, phospholipid, and unsaponifiables (Tzen 2021). Of the sesame oil, oleic acid (C18:1) (~40%) and linoleic acid (C18:2) (~46%) are the two main unsaturated fatty acids and reach up to 85%. The other polyunsaturated fatty acid linolenic acid (C18:3) (~0.4%) and saturated fatty acids including palmitic acid (C16:0) (~8%), stearic acid (C18:0) (~6%), arachidic acid (C20:0) (~0.4%), and behenic acid (C22:0) (~0.6%), comprise about 15% (Anilakumar et al. 2010; Zhang et al. 2019; Langyan et al. 2022). Accorded with the recommendation for low saturated fatty acid level for human health by the World Health Organization (WHO), sesame oil could be regarded as an ideal vegetable oil. The rich unsaturated fatty acids in sesame oil can help prevent cardiovascular diseases, such as arteriosclerosis, by lowering cholesterol and elevating the level of high-density lipoprotein in human blood (Langyan et al. 2022). Of the unsaturated fatty acids, linoleic acid (C18:2) and α-linolenic acid (C18:3) are the essential fatty acids for human body, but cannot be synthesized by itself. Linoleic acid (C18:2) has been proved to inhibit malignant melanoma growth (Smith and Salemo 1992), while α-linolenic acid (C18:3) and its metabolites play vital roles in the visual and brain development in infants and have beneficial effects on suppressing Alzheimer's disease in the aged (Hashimoto and Hossain 2011). In addition, oleic acid (C18:1) with the presence of natural antioxidants in sesame seeds characterized high oxidative stability of the oil. In recent years, breeding new varieties with the high oleic acid (C18:1) content have been a key objective for sesame.

2.2 Sesame Seed Protein and Amino Acid Composition

Sesame seed protein mainly comprises insoluble 11S globulin (α-globulin) (60–70%) and soluble 2S albumin (β-globulin), as well as minor 7S globulin.

The top two protein fractions account for 80–90% of the total protein. In sesame, 11S globulin is the most abundant seed storage protein and presents as a hexamer (300–350 kDa) formed by random assembly of several 11S globulin isoforms (Plietz et al. 1986). Compared with that of soybean, 11S globulin of sesame has the similar subunit structure, but shows more hydrophobic properties. Thus, the solubility, gelation properties, and emulsification capacities of sesame protein are diverse from other crops during food processing (Tzen 2021).

Sesame protein contains 18 types of amino acids, and glutamic acid (4.21%) and arginine (2.79%) are the two main amino acids (Zhang et al. 2021a). Of these, the essential amino acids for human health comprise 31.4% content of total amino acids. Except for lysine, the content of other amino acids in sesame is close to or exceeds the values of high-quality protein from beef and milk recommended for dietary requirement by the Food and Agriculture Organization (FAO) and WHO. Particularly, sesame contains high content of sulfur-containing amino acids, methionine, and cysteine, all of which are attributed mainly to 2S albumin and partly to 11S globulin (Tzen 2021). Thus, the nutritive compositions of sesame protein exhibit the wide usage as food additives for increasing the nutritional quality and food product price. However, the intrinsic property of sesame protein, mainly high hydrophobicity of 11S globulin, does not favor fluid and beverage processing (Anilakumar et al. 2010). After oil crushing, most sesame proteins and cakes are directly processed as animal food or agricultural fertilizer. New more processing techniques should be innovated to increase the nutritional value of sesame protein source.

2.3 Antioxidants

As a specific oilseed crop, sesame is not only used traditionally as a health food, but also as a medicine to prevent diseases. In ancient Chinese pharmacopeia, black sesame seeds were recorded as the medicine and tonic. In Ayurvedic of India, sesame oil was used for external administrations to provide health benefits and treat specific indications (Lahorkar et al. 2009). Current studies point out that the health-promoting properties of sesame are mainly attributed to the antioxidant activity of several bioactive components such as tocopherols and lignans (Andargie et al. 2021; Zhang et al. 2021a).

Tocopherol is one of the important lipid-soluble antioxidants and is mainly present in vegetable oils. Since its discovery as an essential nutrient for reproduction in rats in 1922, tocopherol has been demonstrated to have the effects on scavenging free radicals generated during lipid peroxidation and other radiation-induced oxidation, and thus play a role in preventing chronic diseases such as cardiovascular disease, arteriosclerosis, and cancer (Brigelius-Flohé and Traber 1999; Herrera and Barbas 2001). Tocopherol has four derivative forms, that is, α, β, γ, and δ. Of these, α-tocopherol is the predominant form in nature and presents the highest biological activity by fetal resorption assays (Brigelius-Flohé and Traber 1999). In sesame oil, the content of tocopherol reaches 0.45 mg/g, with 90.5% γ-tocopherol and traces of α-tocopherol and δ-tocopherol (Langyan et al. 2022). In contrast to α-tocopherol

with the oxygen radical scavenging, γ-tocopherol may prevent peroxynitrite-dependent lipid, protein, and DNA damage in the cell (Brigelius-Flohé and Traber 1999). Further reports also showed that γ-tocopherol possesses anti-inflammatory effect by inhibiting cyclooxygenase activity but independent of oxidative activity (Jiang et al. 2000). Moreover, γ-tocopherol and lignans have been found to act synergistically by the intake of sesame seeds to enhance vitamin E activity, which is related to its potential protective effect against cardiovascular disease and cancer (Yamashita et al. 1992).

Lignans are plant phenolic metabolites formed from two molecules of phenylpropanoids. In sesame, lignans are comprised of sesamin, sesamolin, sesaminol, sesamolinol, pinoresinol, sesamol, lariciresinol, matairesinol, piperitol, episesamin, episesaminone, and samin (Andargie et al. 2021). Sesamin and sesamolin are the major lignan components. The contents of that in sesame seeds vary significantly among varieties of sesame accessions (Andargie et al. 2021; Zhang et al. 2021a; Langyan et al. 2022) (Table 2). Sesamin and sesamolin belong to the furofuran family and present in sesame oil, while other minor lignans are grouped into furofuran, tetrahydrofuran, and butyrolactone class. Pinoresinol and sesaminol are mainly present in mono-, di-, and triglycosylated forms. Mono glycosylated lignans are oil-soluble, while diglucoside and triglucoside isoforms are water-soluble.3.

In sesame, the biosynthesis process of lignans, including (+)-sesamin, (+)-sesamolin, and (+)-sesaminol, has been well studied (Fig. 1). Two molecules of coniferyl alcohol from phenylpropanoid metabolism pathway are initially converted to (+)-pinoresinol (a central precursor of lignans) by stereo-selective radical coupling reaction. (+)-sesamin is subsequently formed through sequential oxygenation of (+)-pinoresinol by a cytochrome P450 monooxygenase CYP81Q1 (Ono et al. 2006). Further, CYP92B14, another P450 enzyme, converts (+)-sesamin into (+)-sesamolin and (+)-sesaminol through oxidative rearrangement of α-oxy-substituted aryl groups (Murata et al. 2017). In contrast to the native lignans in sesame seeds, several specific lignans form during seed roasting and sesame oil refining and bleaching. For instance, sesamol is produced from sesamolin by hydrolysis under heating and bleaching with acidic clay, thus is found to have a relatively high content in roasted seed oil but not in unroasted seeds and raw oil (Andargie et al. 2021). The bleaching process converts sesamolin into sesaminol, and also catalyzes the epimerization of sesamin to form episesamin (Fukuda et al. 1986). Differing from sesamin and sesamolin with no or weak antioxidative activity, the converted sesaminol and sesamol are potent antioxidants due to their free phenolic groups (Wan et al. 2015). Therefore, it is considered that the content of antioxidative lignans in sesame products could be raised by the processing such as roasting.

Numerous reports have demonstrated that lignan-rich sesame products as functional nutrients play important roles in preventing diseases. Apart from fatty acids and tocopherols, the sesame lignans, mainly sesamin, sesaminol, sesamol, and sesaminol, attract the most interest of nutritionists, health professionals, even pharmacologists due to their in vitro and in vivo biological activities and therapeutic benefits, including antioxidative, anticancer, anti-inflammatory, anti-hypercholesterolemic, neuroprotective, hepatoprotective, and cardioprotective effects (Table 2).

Table 2 Physical characteristics, content, and main biological functions of sesame lignans, sesamin, sesamolin, sesamol, and sesaminol

Sesame lignan	Molecular formula	Molecule size	Content in seeds (mg/100 g)	Biological function
Sesamin	$C_{20}H_{18}O_6$	354.35	20–1060	Antioxidative and radical scavenging activity in vivo Increasing γ-tocopherol levels in plasma and liver Decreasing blood glucose level and improving body weight and blood pressure Attenuated endothelial dysfunction correlated with cardiovascular diseases Lowering the serum lipid Acting with α-tocopherol synergistically to lower cholesterol level Exerting the antinociceptive and anti-inflammatory effects Protecting kidney against lipid peroxidation and prevent inflammation Protecting liver against ethanol- and carbon tetrachloride-induced damage Suppressing inflammation-induced neurodegeneration Preventing apoptotic cell death and microglial activation Alleviating depression-like behavior and memory deficits, and suppressing age-related cognitive decline Preventing ischemic stroke Extending the mean lifespan Suppressing age-related disorders of kidney Presenting anticancer activity and inhibiting the proliferation of cancer cells
Sesamolin	$C_{20}H_{18}O_7$	370.35	2.4–752	Antioxidative activity in vivo Alleviating the effect of cerebral ischemia by a mixture of sesamin and sesamolin Prolonging median and mean lifespan, and preventing Alzheimer's disease (sesamin, sesamolin, and sesamol synergistically acting) Inducing apoptosis in lymphoid leukemia cells and colorectal cancer cells, and inhibiting growth of leukemia cells Inhibiting premalignant lesions of colon

(continued)

Table 2 (continued)

Sesame lignan	Molecular formula	Molecule size	Content in seeds (mg/100 g)	Biological function
Sesamol	$C_7H_6O_3$	138.12	0–30	Antioxidative activity both in vivo and in vitro Inhibiting lipid peroxidation and mutagenicity of reactive oxygen species Inhibiting myeloperoxidase Reducing blood pressure, preventing vascular dysfunction and exerting both cardioprotective and hepatoprotective effects Accelerating wound healing, and exerting the antinociceptive and anti-inflammatory effects Protecting kidney against lipid peroxidation and prevent inflammation Presenting neuroprotective and myoprotective effect Improving depression and memory deficits Antiaging effects Presenting anticancer activity and inhibiting the proliferation of cancer cells
Sesaminol	$C_{20}H_{18}O_7$	370.40	0–1.4	Antioxidative activity both in vivo and in vitro Stimulating elastin and collagen production in the skin, and accelerating wound healing Protecting against neuronal apoptosis Protecting skin against photodamage caused by chronic UV exposure Stimulating collagen and elastin production and migration of keratinocytes Inhibited cancer cells in vitro

Source: Andargie et al. (2021), Zhang et al. (2021a), and Langyan et al. (2022)

Major sesame lignans possess both or either in vitro and in vivo antioxidant activity, which accounts for a large part of the health benefits of sesame consumption (Ray and Katyal 2016). For example, sesamin has been reported to protect liver from oxidative damage in vivo, which results from its metabolites with the catechol moieties presenting radical scavenging and antioxidant activities (Nakai et al. 2003). Similar to sesamin, sesamolin has no antioxidant activity in vitro, but presents high superoxide scavenging effects in vivo. After converted from sesamolin, sesamol possesses both in vitro and in vivo antioxidant activity, thus has the effects in inhibiting lipid peroxidation, and mutagenicity of reactive oxygen species and hydroxyl radical-induced DNA damage (Joshi et al. 2005). In addition, sesamin

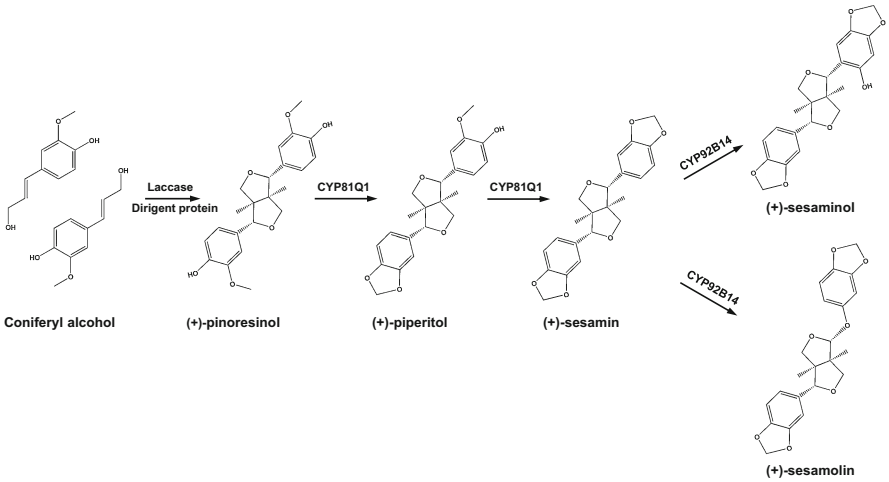


Fig. 1 The biosynthesis pathway of lignans in sesame from (+)-pinoresinol to (+)-sesamolin and (+)-sesaminol

exerts indirectly antioxidant effects via the inhibition of the catabolism of tocopherol (Parker et al. 2000). The effect of the intake of sesamin actually increases the level of tocopherol in vivo, which is the physiological antioxidant mentioned above.

Some reports indicated that sesame lignans were likely responsible for prevention and treatment of neurodegenerative diseases such as Parkinson's and Alzheimer's disease and multiple sclerosis (Ray and Katyal 2016), which were in part attributed to oxidative damage. Amelioration of oxidative stress by supplying sesame lignans was demonstrated to delay the aging of aerobic organisms (Zuo et al. 2013). Sesame lignans also have the effects on prevention of cardiovascular diseases by lowering the levels of lipogenesis, blood pressure, and cholesterol (Andargie et al. 2021). In addition, sesame seed and sesame oil were found to possess wound-healing activity, while sesamin and sesamol were accounted for a part of the effect (Shenoy et al. 2011; Monteiro et al., 2014). Remarkably, the anticancer activity of sesame lignans, mainly sesamin and sesamol, has been extensively tested in tumor cell lines and animals, and the positive effects were acted to inhibit the proliferation of lung, prostate, breast, cervical, colon, liver cancer, skin papillomas, and leukemia cells (Langyan et al. 2022). Recently, the therapeutic benefits of sesame lignans have been comprehensively reviewed in detail (Andargie et al. 2021; Langyan et al. 2022). We expect sesame lignans to be applied for further pharmacological and clinical studies with regard to their physicochemical properties and biological activities.

3 Genetic Resources and Nutrition-Related Traits

According to data statistics, about 20,000 sesame germplasm resources from the world are collected and reserved in sesame germplasm reservoirs mainly in South Korea, India, China, the United States, and some African countries now. Substantial

evidence indicated that the content range of oil, protein, and other nutrients in sesame seeds is evidently huge among the diverse germplasm accessions and exhibits the contribution of both genotype and environment (Tashiro et al. 1990; Li et al. 2014). For hundreds of the world sesame germplasm, the content of oil and protein ranges from 27.9% to 59.8% and 16.7% to 31%, respectively (Li et al. 2014).

To rapidly and concisely evaluate the content of sesame nutrients in seeds, the near-infrared reflectance spectroscopy (NIRS) method has been applied to establish evaluation standard curves and evaluate the contents of oil, protein, fatty acid components, lignan, edible fiber, and polysaccharide, respectively, in sesame seeds (Yuan et al. 2022). With the aid of evaluation platform, Sesame Research Center, Henan Academy of Agricultural Sciences (HRSC, HAAS) evaluated the content of oil, protein, the main fatty acid components of the 857 sesame germplasm planted in Hainan, Henan, and Yuanmou of China in 2019 (Table 3) (unpublished data). Evident variation is presented in oil and protein content. Meanwhile, the contents of palmitic acid (C16:0), palmitoleic acid (C16:1), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), and linolenic acid (C18:3) also fluctuated with genotypes and environments. The assayed results of 126 sesame germplasm by the NIRS method showed that their content of oleic acid, linoleic acid, palmitic acid, and stearic acid ranged from 35.1–51.8%, 31.6–47.8%, 7.2–11.3%, and 4.4–7.7%, respectively (Yuan et al. 2022).

As regards the contents of antioxidant, Tashiro et al. (1990) analyzed the 42 sesame accessions from Japan and detected that the contents of sesamin and sesamolin varied from 0.07 to 0.61 mg/g oil and 0.02 to 0.48 mg/g oil, respectively. Rangkadilok et al. (2010) investigated the sesame accessions from Thailand and found that the mean content of sesamin and sesamolin was 1.55 mg/g and 0.62 mg/g seed, respectively, and the total tocopherols was 50.9–211 μg/g seed. While the ranges of sesamin, sesamolin, and tocopherol contents were 0.93–2.89 mg/g, 0.30–0.74 mg/g, and 0.30–0.65 mg/g in commercial sesame oils. The extensive variability in sesamin, sesamolin, and tocopherol contents demonstrated the nutrition variation of different sesame products. Wang et al. (2013) analyzed sesamin and sesamolin contents of the 62 sesame cultivars from China using high-performance liquid chromatography (HPLC), which showed the sesamin and sesamolin contents ranged from 0.82 to11.05 mg/g and 1.35 to 6.96 mg/g, respectively, and the total sum of sesamin and sesamolin among about 60.0% of the assayed Chinese cultivars varied from 6.0 to 9.0 mg/g. Moreover, some Chinese landraces possessed significantly higher sesamin and sesamolin contents than those of cultivars. The concise phenotypes of sesame nutrients with highly efficient evaluation techniques supply the basis for genetic mechanism analysis of sesame nutrients and application in sesame production.

4 Genetic Diversity Assessment of Sesame Germplasm with Nutrition-Related Traits

Genetic diversity indicates the gene or genome sequence variation of the individuals of a species, which is believed as a key aspect to reflect the evolution of a species and the adaption to environmental stresses. Most of the biological diversity studies in plants focus on both phenotypes and genome variation.

Table 3 Variation of nutrition components of the 857 sesame germplasm under different environments

Seed nutrition	Cultivated location	Maximum content (%)	Minimum content (%)	Average value (%)
Oil	Sanya	61.60	39.73	51.37
	Yuanmou	58.53	40.72	49.47
	Yuanyang	56.84	42.59	50.62
	Average	58.99	41.01	50.49
Protein	Sanya	27.79	15.48	22.21
	Yuanmou	26.73	16.47	23.55
	Yuanyang	27.48	16.43	22.26
	Average	27.33	16.13	22.67
Palmitic acid (C16:0)	Sanya	10.68	2.81	8.03
	Yuanmou	10.09	3.28	7.83
	Yuanyang	15.23	4.66	7.79
	Average	12.00	3.58	7.88
Palmitoleic acid (C16:1)	Sanya	0.16	0.01	0.11
	Yuanmou	0.18	0.01	0.11
	Yuanyang	0.30	0.04	0.12
	Average	0.21	0.01	0.11
Stearic acid (C18:0)	Sanya	7.09	4.33	5.57
	Yuanmou	8.71	4.08	6.35
	Yuanyang	7.86	4.16	5.86
	Average	7.89	4.19	5.93
Oleic acid (C18:1)	Sanya	62.98	32.33	44.20
	Yuanmou	65.45	40.13	49.83
	Yuanyang	57.17	24.24	44.88
	Average	61.87	32.23	46.30
Linoleic acid (C18:2)	Sanya	53.09	32.77	41.40
	Yuanmou	43.86	22.86	35.05
	Yuanyang	51.57	29.70	41.01
	Average	49.51	28.44	39.15
Linolenic acid (C18:3)	Sanya	0.45	0.15	0.28
	Yuanmou	0.52	0.16	0.31
	Yuanyang	0.49	0.18	0.29
	Average	0.49	0.16	0.29
Arachidic acid (C20:0)	Sanya	0.88	0.37	0.63
	Yuanmou	0.93	0.33	0.65
	Yuanyang	0.89	0.38	0.60
	Average	0.90	0.36	0.63

Up to now, extensive studies on the genetic diversity of sesame germplasm have been performed using various molecular markers, such as AFLP, SRAP, ISSR, SSR, EST-SSR, SNPs, and InDels (Wei et al. 2008; Zhang et al. 2010; Zhang et al. 2012; Wu et al. 2014; Wei et al. 2015; Cui et al. 2017; Yadav et al. 2022). Using SSR markers, a genetic diversity survey was conducted in sesame germplasm panel consisting of 545 accessions, including 390 lines from China, 149 lines from 19 other countries, and 6 sesame wild relatives (Yue et al. 2012). As a result, a total of 106 alleles were detected in the germplasm lines. The three major clusters were grouped in an UPGMA tree. The first cluster included 4 sesame wild relatives, and the second cluster consisted of 28 lines that were mainly from Ethiopia, while the third cluster contained the rest of 513 lines. In addition, the genetic diversity study was also conducted with the 366 elite germplasm lines using specific-locus amplified fragment sequencing (SLAF-seq) data (Cui et al. 2017). Of these, 329 lines were from 18 provinces in China, and the rest 37 lines were from 11 other countries. Based on the population structure analysis through STRUCTURE program, 144 and 111 germplasm lines were assigned to the two subgroups, respectively, and the rest 111 lines were in a mixed subgroup. For all of the lines, the sequence diversity (π) of the panel was estimated as 1.1×10^{-3}. The estimates of sequence diversity were 1.0×10^{-4}, 2.7×10^{-4}, and 3.6×10^{-4} for subgroups 1, 2, and the mixed, respectively (Cui et al. 2017).

The large genetic diversity of germplasm means the diversity of various nutrition-related traits in sesame. In fact, nutrition-related traits such as the contents of oil, protein, and lignans vary significantly in sesame germplasm, which provides a foundation for subsequent genetic research on sesame nutrition-related traits. On the other hand, the diversity studies of sesame germplasm with nutrition-related traits showed that the diversity of nutrition-related traits from the same germplasm subgroup could not be clearly distinguished from other subgroups, that is, the germplasm from the same subgroup also contain highly variable nutrition-related traits. For example, in a panel containing the 369 core sesame germplasm, both oil and protein contents significantly varied among the germplasm, reflecting a broad phenotypic diversity and showing no significant correlation with the group structure of germplasm (Li et al. 2014). Thus, further fully exploring genetic variation should be indispensable for applying germplasm and conducting breeding programs.

5 Classical Genetics of Nutrition-Related Traits

Sesame is the sole cultivated species in *Sesamum*. Digging new germplasm and creating new materials with elite traits and genotypes are urgent for improving the sesame genetics and breeding research (Ju et al. 2021). Most nutrition-related traits belong to complex quantitative traits in sesame and other oilseed crops, being affected by both genotype (G) and environment (E). Li et al. (2014) systematically investigated the phenotype variation in a panel of the 369 sesame germplasm and indicated that the oil and protein contents were affected by G and G $\times$ E, and the oil content is negatively correlated with protein content (Wei et al. 2013; Li et al. 2014).

Moreover, the broad-sense heritability value (H^2) of oil content trait was higher than that of protein content, which suggested that the variation of oil content is mainly affected by genotype. For sesame oil content trait, many researchers conduct the classic genetics analysis to explain its genetic variance, and gene actions of the additive, nonadditive types or both interaction were found in different crossing populations (Wei et al. 2021). It meant that the advanced generation selection could not be efficient to improve oil content in sesame variety breeding. Besides, the specific combining ability variance of sesame oil and protein contents was respectively in moderate and high proportions. Especially, there is almost no heterosis in oil content, while protein content showed significant negative heterosis (Murty 1975; Sumathi and Muralidharan 2014).

As regards the fatty acid components, few genetic studies have been performed on sesame because of deficiency of specific germplasm and phenotype data in the early stage. Recently, the genetic character analysis of oleic acid content was performed with the aid of a mutant with a high oleic acid content. The first mutant with a high oleic acid content, named HO995, was created using EMS mutagenesis by HRSC, HAAS in 2017. Its oleic acid content increases up to more than 75% and is significantly higher than the average value of control samples (40%). Genetic background analysis indicated that the oleic acid content for the mutant HO995 is controlled by one incomplete gene pair (unpublished data). Based on the mutant and hybrid breeding technique, Yuzhi HO995, the first sesame variety with a high oleic acid content in the world, was released in China in 2022.

Sesamin and sesamolin contents were significantly and positively correlated with each other (Xu et al. 2020). Their contents are quantitative traits in sesame germplasm accessions and are controlled by a common polygenic system (Usman et al. 2020). However, the classic genetic analysis determined that the sesamin and sesamolin contents fit to an additive-dominant model, and the additive variance exceeded the dominance variance in both contents (Ogata and Kato 2016). In addition, by analyzing the sesamin and sesamolin contents in the 18 sesame cross-combinations, the heterosis of the sesamin and sesamolin contents was found to be mainly mid-parent and mostly negative, while only 5 of 18 combinations present over-high parent heterosis (Xu et al. 2020). Thus, the selection and combination of elite parents should be critical to breed new sesame varieties with high sesamin and sesamolin contents.

Not only significant correlation was determined between lignan contents, but also between lignan contents and other nutrition-related traits. Wu et al. (2017) found the significant positive correlation between the oil and sesamin contents in sesame seeds ($r = 0.608$, $P < 1\%$), but no correlation between the oil and sesamolin contents. In addition, the protein content was considered to be negatively correlated with the sesamin content as the oil content was negatively correlated with protein content. Positive correlation among multiple nutrition-related traits is usually beneficial for breeding high-quality sesame varieties, but the adverse effects of negative correlation should also be carefully considered during the breeding process.

6 QTL/QTN Mapping of Nutrition-Related Traits

To decipher the genetic basis of the seed quality traits in sesame, so many studies of molecular markers discovery, genetic map construction, population association analysis, and quantitative trait locus (QTL) mapping have been performed in the past two decades (Table 4). Wei et al. (2013) analyzed the oil, protein, oleic acid, and linoleic acid content variation of the 216 Chinese sesame accessions using 79 molecular markers. Only marker M15E10–3 was confirmed to associate with the oil

Table 4 Genetic and association mapping of nutrition-related components in sesame seeds

Traits	Marker type	Marker number	Population	Major result	Reference
Oil, protein, oleic acid, and linoleic acid	SRAPs, SSR, AFLP	79	216 germplasm lines from China	Ten markers associated with oil, protein, oleic acid, and linoleic acid concentration were identified	Wei et al. 2013
Oil and protein	SSR	112	369 worldwide germplasm lines	A total of 19 and 24 SSR markers associated with oil and protein contents, respectively, were detected	Li et al. 2014
Mineral-nutrients	SNP, InDel	19,309	149 F_2 lines	A bunch of QTLs significantly associated with Zn, Fe, Cu, Mn, Ca, Mg, K, P, and S concentration were detected	Teboul et al. 2020
Sesamin and sesamolin	SSR	424	548 RILs (F_8)	Sixteen and 10 QTLs associated with sesamin and sesamolin contents, respectively, were detected	Xu et al. 2021
Vitamin E	SNP, InDel	5962	96 core germplasm accessions	LG08_6621957 associated with γ-tocopherol significantly; SLG03_13104062 weakly associated with β-tocotrienol	He et al. 2019
Oil, protein, sesamin, sesamolin, fatty acid, and unsaturated fatty acid	SNP	1,805,413	705 worldwide germplasm lines	Total of 549 associated loci for 56 agronomic traits	Wei et al. 2015

content under two environments. Li et al. (2014) performed an association analysis of oil and protein contents of the 369 worldwide sesame lines using 112 polymorphic SSR markers under multienvironments. A total of 19 and 24 SSR markers were found significantly associated with oil and protein content with the R^2 values of 4–29% and 3–29%, respectively.

In order to mine QTLs associated with the content of mineral-nutrient elements, Teboul et al. (2020) conducted genetic mapping of micronutrient (zinc, Zn; iron, Fe; copper, Cu; and manganese, Mn) and macronutrient (calcium, Ca; magnesium, Mg; potassium, K; phosphorus, P; and sulfur, S) contents with the 149 lines of an F_2 population. Of the mineral elements, Fe concentration was positively correlated with those of Zn, Cu, and Mg. P content was positively correlated with those of Zn, Fe, Cu, Mg, and S, rather than with those of Mn, Ca, and K. A bunch of QTLs significantly associated with Zn, Fe, Cu, Mn, Ca, Mg, K, P, and S were detected, with the explanation of the variance of 7.2–45.7%, 6.1–21.6%, 6.3–30.4%, 7.1–22.2%, 6.1–21.4%, 6.8–12.0%, 7.5–10.7%, 5.9–7.8%, and 7.2–12.5%, respectively.

The sesamin and sesamolin contents vary broadly in sesame germplasm. Some reports reflected that sesamin and sesamolin contents are quantitative traits controlled by polygenes in sesame (Wei et al., 2015; Xu et al., 2021). In order to dig the candidate genes and the potential networks regulated the sesamin and sesamolin content in sesame, Xu et al. (2021) firstly constructed a genetic map with the 548 recombinant inbred lines (RILs, F_8) derived from a cross between ZZM2748 (male parent) and Zhongzhi 13 (female parent). A total of 424 SSR markers distributed in the 13 linkage groups of the genetic map with a total length of 1869.78 cM. As a result, 16 and 10 QTLs associated with sesamin and sesamolin contents, respectively, were detected, with the phenotypic variation explanation (PVE) of 1.15–67.69% and 1.87–46.05%, respectively. Of these, qSmin_3.1, qSmin_9.2, qSmin_11.1, and qSmol_11.1 were detected in the two environments. The loci of qSmin11–1 and qSmol11–1 were detected in the same region, explaining 67.69% and 46.05% of the phenotype variation of sesamin and sesamolin, respectively. Further fine mapping of the loci qSmin11–1 and qSmol11–1 to the reference genome determined that two genes, *SIN_1005755* and *SIN_1005756*, were proposed as the candidate gene involved in sesamin and sesamolin biosynthesis. Meanwhile, He et al. (2019) conducted a GWAS for vitamin E content, composed of β-tocopherol, γ-tocopherol, and β-tocotrienol, using 5962 genome-wide markers acquired from 96 core sesame accessions. Locus LG08_6621957 detected was significantly associated with γ-tocopherol content. The allelic variation of base "G" to "A" resulted in the increase of the average γ-tocopherol content in sesame seeds.

Remarkably, a comprehensive GWAS was executed to reveal the variants associated with the nutrient components, including oil, protein, sesamin, sesamolin, fatty acid, unsaturated fatty acid, and the ratio fatty acid/unsaturated fatty acid among the 705 sesame accessions (Wei et al. 2015). The validation of quantitative trait nucleotides (QTNs) indicated six candidate genes were for seed oil content (SIN_1003248, SIN_1013005, SIN_1019167, SIN_1009923, *SiPPO*/SIN_1016759 and *SiNST1*/SIN_1005755), while

one candidate gene was associated with protein content (*SiPPO*/SIN_1016759). Meanwhile, eight candidate genes (i.e., *SiKASI*/SIN_1001803, *SiKASII*/SIN_1024652, *SiACNA*/SIN_1005440, *SiDGAT2*/SIN_1019256, *SiFATA*/SIN_1024296, *SiFATB*/SIN_1022133, *SiSAD*/SIN_1008977 and *SiFAD2*/SIN_1009785) were for the fatty acid compositions, and one gene *SiNST1*/SIN_1005755 for sesamin and sesamolin content. All the above information provides useful molecular markers and candidate genes for further genetics and breeding research of sesame seed quality.

7 Discovery of Nutrition-Related Genes

The key genes regulating the fatty acid biosynthesis and metabolism in sesame and other oilseed crops always attract more attention. For sesame, cDNA library screening, homolog gene cloning, fine linkage mapping, genome-wide association studies, and comparative genomic and transcriptomic analysis have been applied to dig functional genes-related nutrient traits (Table 5). With the aid of genomic resources, more than a dozen functional genes related to the oil content and antioxidant biosynthesis have been identified in sesame. To reveal its genetic basis of high oil content, the FA gene families and comparative genomic analysis were executed between sesame and other oilseed crops (Miao et al. 2021b). Sesame has unexpectedly lower gene copy numbers (708) of the lipid-related genes compared to soybean (1298), five lipid-related gene families were found to significantly expand and two lipid degradation-related families were contracted in sesame (Wang et al., 2014). Among these families, the transfer protein type 1 (LTP1) genes had the functional activities in lipid biosynthesis according to transcriptome analyses, which suggested their potential role in high oil accumulation in sesame seeds (Wang et al., 2014). Furthermore, genome-wide sequence analysis indicated that two genes *SiLTPI.23* and *SiLTP1.28* integrated the *SiLTPs* expression profiles and were related to the variation of oil content in sesame varieties. Ectopic expression of *SiLTP1.23* in *Arabidopsis thaliana* proved that the gene played the important role in significant increase of oil content and regulation of the ratio of fatty acid compositions in developing seeds (Song et al. 2021). In addition, three genes, SINPZ1100015 (NAC domain-containing protein 43), SINPZ1201700 (trehalose-phosphate synthase I, *SiTPS1*), and SINPZ1201748 (3-oxoacyl-[acyl-carrier-protein] synthase I, *SiKAS1*), were significantly associated with the oil and fatty acid contents under all three environments by GWAS (Zhou et al. 2022). For SINPZ1100015, an allele variation of "C" to "A" resulted in the significant increase of the palmitic acid (C16:0) and linolenic acid (C18:3) contents, but decrease of the oil content in sesame seeds. For *SiTPS1* and *SiKAS1*, the allelic SNP variations from the reference base "A" to "T" and "C" to "A," respectively, were related to the reduction in the palmitic acid (C16:0) content. The above three genes were significantly and highly expressed during the seed development in sesame varieties with high oil content. Overexpression of *SiTPS1*in *Arabidopsis* also exhibited a significant increase in the oil, palmitic acid (C16:0) and oleic acid (C18:1) contents, and the reduction in the

Table 5 Genes related to the oil content and lignan biosynthesis in sesame

Gene	Protein	Function	Trait	References
SiLTP1.23	Nonspecific lipid-transfer protein	Increasing the oil content and regulating fatty acid composition metabolism	Oil biosynthesis	Song et al. 2021
SiLTP1.28	Nonspecific lipid-transfer protein	Involved in fatty acid metabolism	Oil biosynthesis	
SINPZ1100015	NAC domain-containing protein 43	Involved in the fatty acid metabolism	Oil biosynthesis	Zhou et al. 2022
SINPZ1201700/ *SiTPS1*	Trehalose-phosphate synthase I	Maintains fatty acid composition through palmitic acid, oleic acid, and linoleic acid metabolism	Oil biosynthesis	
SINPZ1201748/ *SiKAS1*	3-Oxoacyl-[acyl-carrier-protein] synthase I	Maintaining fatty acid composition through palmitic acid biosynthesis	Oil biosynthesis	
SeFAD2	Oleic acid desaturase	Catalyzing desaturation of oleic acid to linoleic acid	Fatty acid biosynthesis	Jin et al. 2001
CYP81Q1	Cytochrome P450 monooxygenase/ piperitol/sesamin synthase	Converting (+)-pinoresinol to (+)-sesamin through two consecutive oxygenation	Sesamin biosynthesis	Ono et al. 2006
CYP92B14	Cytochrome P450 monooxygenase	Converting (+)-sesamin to (+)-sesamolin and (+)-sesaminol	Sesamolin biosynthesis	Murata et al. 2017

UGT71A9	UDP-sugar-dependent glucosyltransferases	Glucosylating at the 2-hydroxyl group of (+)-sesaminol to produce sesaminol monoglucoside	Sesaminol glucosylation	Noguchi et al. 2008
UGT94D1	UDP-sugar-dependent glucosyltransferases	Catalyzing β1 → 6-O-glucosylation of sesaminol glucosides and preferring sesaminol monoglucoside as a substrate	Sesaminol glucosylation	
UGT94AG1	UDP-sugar-dependent glucosyltransferases	Catalyzing the β1 → 2-O-glucosylation of sesaminol glucosides to form sesaminol triglucoside	Sesaminol glucosylation	Ono et al. 2020
UGT94AA2	UDP-sugar-dependent glucosyltransferases	Catalyzing β1 → 6-O-glucosylation of sesaminol glucosides and preferring (+)-sesaminol 2-O-β-d-glucosyl-(1 → 2)-O-β-d-glucoside as a substrate	Sesaminol glucosylation	
LOC105172736/ *SinPRL2*	Bifunctional pinoresinol-lariciresinol reductase	Converting pinoresinol to lariciresinol and subsequently lariciresinol to secoisolariciresinol	Lariciresinol and secoisolariciresinol biosynthesis	Andargie et al. 2021
SiWRKY67	WRKY transcription factor	Regulating melatonin content	Melatonin biosynthesis	Wang et al. 2022

linoleic acid (C18:2) content. Thus, *SiTPS1* was considered a key regulator of fatty acid biosynthesis and triacylglycerols formation in sesame (Zhou et al. 2022).

In plants, *FAD2* and *FAD6* function respectively for endoplasmic reticulum- and plastid-derived ω-6 fatty acid desaturases (FAD), which catalyze desaturation of oleic acid to linoleic acid (C18:2). The sesame *SeFAD2/FAD2*, essential for the biosynthesis of polyunsaturated fatty acids, was obtained from the cDNA library by the homolog sequence alignment and synthesized in developing seeds using RT-PCR (Jin et al. 2001). *SeFAD2* transcripts were regulated by an intron-mediated regulatory mechanism. Some reports showed that the other two genes *SebHLH* and *SeCKI* enhanced the expression of *SeFAD2* via the *SeCKI-SebHLH-SeFAD2* pathway, in which *SeCKI* phosphorylated *SebHLH* transcription factor, and phosphorylated *SebHLH* subsequently mediated trans-activation of the *SeFAD2* promoter (Kim et al. 2007; Kim et al. 2010). These lipid-related genes are the essential resources for genetic improvement of high oil content and suitable fatty acid composition in sesame.

Besides the oil content, the variation of antioxidants content in sesame attracts more attention for the contribution to oil quality and health-promoting properties. Till now, the biosynthesis pathways of sesame lignan have been comprehensively reviewed (Andargie et al. 2021). Several genes, including *CYP81Q1*, *CYP92B14,* and UDP-sugar-dependent glucosyltransferases (UGT), have been identified to involve in the metabolic pathway of sesame lignans (Table 5). *CYP81Q1* encoding a cytochrome P450 monooxygenase was screened from a sesame seed cDNA library with probe mixtures of cytochrome P450 genes. The consistent expression patterns were correlated with the accumulation pattern of (+)-sesamin during seed development. Further studies demonstrated that *CYP81Q1* played a role as a piperitol/sesamin synthase in the conversion of (+)-pinoresinol to (+)-sesamin through two consecutive oxygenation processes (Ono et al. 2006). Another P450 monooxygenase gene *CYP92B14* was located in a RAD-seq genetic map with a sesame RIL lines (F_6 generation) derived from the cross of a (+)-sesamolin-deficient accession #4294 and a high (+)-sesamolin content cultivar ITCFA2002 (Murata et al. 2017). The results proved that CYP92B14 was responsible for the conversion of (+)-sesamin to (+)-sesamolin and (+)-sesaminol through oxidative rearrangement of α-oxy-substituted aryl groups. In the allele of *CYP92B14* from #4294, the insertion of "T" resulted in the lack of four C-terminal amino acids and finally led to the deficiency of (+)-sesamolin. Differing from the function of *CYP92B14* and *CPR1*, co-expression of *CYP81Q1* with *CYP92B14* and *CPR1* in yeast increased the amount of (+)-sesaminol and (+)-sesamolin. The results suggested there exists functional cooperation between *CYP81Q1* and *CYP92B14* (Murata et al. 2017).

In addition, two UGT genes, *UGT71A9* and *UGT94D1*, were identified and screened from a cDNA library according to the lignan glucosylating activity (Noguchi et al. 2008). The other two distinct UGT genes, *UGT94AG1* and *UGT94AA2,* were obtained based on the transcriptomic co-expression and enzymatic characters (Ono et al. 2020). *UGT71A9* glucosylated at the 2-hydroxyl group of (+)-sesaminol to produce sesaminol monoglucoside, while UGT94AG1 catalyzed the β1 → 2-O-glucosylation of sesaminol glucosides to form finally sesaminol

triglucoside. In sesame seeds, *UGT94D1* and *UGT94AA2* showed coordinated spatio-temporal expression pattern and possessed the redundant function in displaying β1 → 6-O-glucosylation activity toward sesaminol glucosides during sesaminol triglucoside biosynthesis. More studies showed that UGT94AA2 might prefer (+)-sesaminol 2-O-β-d-glucosyl-(1 → 2)-O-β-d-glucoside as a substrate, while UGT94D1 preferentially glucosylated sesaminol monoglucoside (Ono et al. 2020). The results further suggested that the β-O-glucosylation of sesaminol triglucoside biosynthesis is involved in β1 → 2-O-glucosylation of sesaminol monoglucoside induced by UGT94AG1 and the subsequent β1 → 6-O-glucosylation by UGT94AA2 or β1 → 6-O-glucosylation of sesaminol monoglucoside by UGT94D1, followed by UGT94AG1-mediated β1 → 2-O-glucosylation (Ono et al. 2020).

Melatonin (N-acetyl-5-methoxytryptamine) is another kind of important antioxidant with health-promoting effects and always ubiquitously exists in plants, mammals, and human beings. Among sesame germplasm resources, the content of melatonin varied significantly from 0.04 to 298.62 ng/g (Wang et al. 2022). Through GWAS analysis, a WRKY transcription factor *SiWRKY67* was screened and associated with melatonin content trait in sesame. The transient expression of *SiWRKY67* was correlated with the melatonin accumulation in *Agrobacterium rhizogenes*-mediated sesame hairy roots (Wang et al. 2022). Gene sequence analysis indicated that there was a nucleotide polymorphism of *SiWRKY67* in various sesame accessions. The molecular marker could be applied for high melatonin variety improvement in sesame.

8 Available Genome Resources and Databases

So far, hundreds of sesame varieties have been bred and released in the major production countries by mainly using the traditional breeding techniques (Zhang et al. 2021c). However, the relatively narrow genetic variation basis of germplasm accessions limits the genetics and breeding for sesame. Improving the sesame varieties with more elite agronomic traits seems more difficult. Impelling the molecular genetics and genomics research becomes the necessary solution to expand the genetic basis of key agronomic traits in sesame. With the development of genome sequencing technologies, the sesame genome project was implemented since 2010 (Zhang et al. 2013). Till now invaluable genomic resources were generated and stimulated the progress of sesame genomics and molecular breeding (Zhang et al. 2013; Wang et al. 2014). Particularly, the release of several genome drafts of sesame leads to sesame breeding into a new era.

The genome size of sesame is estimated at about 369 Mb by the flow cytometry data (Zhang et al. 2013). Sesame is considered a potential model of oilseed crop for genetic studies due to the small genome size with other characters including high production, short life cycle, and high adaptability for harsh environments (Wang et al. 2014; Zhang et al. 2019, 2021a). The first draft genome with 293.7 Mb of size for sesame cultivar 'Yuzhi 11' was sequenced by Solexa platform (Zhang et al. 2013). Subsequently, the draft genome was optimized by assembling "ABI3730xL+

Roche/454+ Illumina/Solexa + PacBio SMRT" sequencing data and assisting with the "SNP genetic map + BAC-FISH physical map + Hi-C library data." The final genome draft of 'Yuzhi 11' (version 3) reached 335.2 Mb in size with the 13 pseudo-chromosomes and contained the estimated 31,462 genes (Miao et al. 2021a). In parallel, another genome draft for sesame cultivar 'Zhongzhi 13' was released with 274 Mb in size, encompassing estimated 27,148 genes distributed on 16 linkage groups (LGs) (Wang et al. 2014). Both genome drafts are being applied as reference sesame genomes for the world researchers for genomics and comparative genomics analysis nowadays. Moreover, the other three genome assemblies from two sesame landraces (Baizhima and Mishuozhima) and a cultivar Swetha were also sequenced for sesame genome assembly, which provided more valuable resources for structural, functional, and comparative genomic analyses for sesame (Wei et al. 2016) (Table 6). Based on the above de novo genome assemblies, sesame pan-genome and comparative genomics analysis indicated that there were 12 genes involved in lipid biosynthesis and metabolism were under positive selection, which might contribute to the high oil content in sesame (Yu et al. 2019). Integrating comparative genomic and transcriptomic data indicated that the high genetic diversity of lipid-related genes might be related to the wide variation of sesame oil content (Wang et al. 2014). In addition, the chromosome-scaled genomes of the six wild *Sesamum* relatives, that is, *S. alatum* ($2n = 2x = 26$, 3651), *S. angustifolium* ($2n = 2x = 32$, G01), *S. angolense* ($2n = 2x = 32$, K16), *S. latifollum* ($2n = 2x = 32$, KEN1), *S. calycinum* ($2n = 2x = 32$, KEN8), and *S. radiatum* ($2n = 4x = 64$, G02), were de novo assembled, respectively (Zhang et al. 2021b). Genomes of the key *Sesamum* species supply the precious genome information for sesame genetics and breeding research in the world (Table 6).

Besides the genome sequence resources, massive genome resequencing data for more than 2000 sesame accessions and population lines were obtained and released in the past decade (Wei et al. 2015; Zhang et al. 2016; unpublished data). Based on the released multitude genome sequencing and resequencing data and transcriptomes data, five comprehensive online databases have been set for sesame and referred to the Sesame Genome Project (http://www.sesamegenome.org), Sinbase 2.0 (updated version of Sinbase, http://www.sesame-bioinfo.org/Sinbase2.0), SesameHapMap (http://202.127.18.228/SesameHapMap), SesameFG (http://www.ncgr.ac.cn/SesameFG), and SisatBase (http://www.sesame-bioinfo.org/SisatBase). The four public databases, viz., NCBI (http://www.ncbi.nlm.nih.gov/genome/?term=sesame), ocsESTdb (http://www.ocri-genomics.org/ocsESTdb/index.html), PTGBase (http://www.ocri-genomics.org/PTGBase/index.html), and PMDBase (http://www.sesame-bioinfo.org/PMDBase), include sesame genomic resources at the same time.

With the aid of sesame genome databases, studies of molecular marker development, genetic diversity and variation analysis, gene fine-mapping, and genome-wide association studies (GWAS) are being developed accordingly (Wei et al. 2015; Dossa et al. 2017; Murata et al. 2017; Zhang et al. 2016). For example, Sinbase 2.0, a versatile web-based database, including genome, comparative genomes, genetic and phenotype information, supplies an important source to study agronomic

Table 6 The genome assemblies for cultivated sesame and wild *Sesamum* relatives

Species name	Sample	Location	GenBank assembly accession	Genome size (Mb)	Organization	Reference
S. indicum	Zhongzhi 13	China	GCA_000512975.1	274	Oil Crops Research Institute, Chinese Academy of Agricultural Sciences	Wang et al. 2014
S. indicum	Baizhima	China	/	267	Oil Crops Research Institute, Chinese Academy of Agricultural Sciences	Wei et al. 2016
S. indicum	Mishuozhima	China	/	254	Oil Crops Research Institute, Chinese Academy of Agricultural Sciences	Wei et al. 2016
S. indicum	Swetha	New Delhi, India	GCA_000975565.1	340.5	National Bureau of Plant Genetic Resources	Yu et al. 2019
S. indicum	Yuzhi 11[a]	China	/	335.2	The Sesame Genome Working Group	Miao et al. 2021a
S. alatum	3651	Sudan	/	528.4	The Sesame Genome Working Group	Zhang et al. 2021b
S. latifollum	KEN1	Kenya	/	369.1	The Sesame Genome Working Group	Zhang et al. 2021b
S. angolense	K16	Kenya	/	313.0	The Sesame Genome Working Group	Zhang et al. 2021b
S. calycinum	KEN8	Kenya	/	300.8	The Sesame Genome Working Group	Zhang et al. 2021b
S. angustifolium	G01	Congo	/	300.7	The Sesame Genome Working Group	Zhang et al. 2021b
S. radiatum	G02	Congo	/	625.0	The Sesame Genome Working Group	Zhang et al. 2021b

[a]Two versions of genome assemblies for sesame cultivar 'Yuzhi 11' have been released in NCBI

traits regulation in sesame in detail (Wang et al. 2021). A fatty acid biosynthesis and metabolism-related gene database of *Sesamum* species has been locally constructed to reveal the genetic nature of lipid biosynthesis in *Sesamum* (Miao et al. 2021b). The completion of the sesame genome project covering sesame and the six wild species has greatly promoted the progress of molecular genetics and breeding in sesame.

9 Molecular Breeding Strategy for the Development of Nutrition-Related Traits in Sesame

In the past decade, the discovery of so many functional QTLs and genes related to important traits in major crops, such as rice, wheat, corn, and soybean, presents progress in molecular genetics and functional genomics analysis and molecular breeding. For the modern breeding strategy, obtaining appropriate breeding elements including QTLs, genes, and molecular markers associated with the key agronomic traits are the predetermined objective (Zhang et al. 2019). Being an orphan crop, the progress in sesame molecular breeding and -omics research gives more expectation for sesame breeding. Breeding superior sesame varieties with high seed quality is a key goal for sesame breeders. Based on the genomics-assisted breeding in sesame, the Sesame Genome Working Group proposed the scheme of the molecular design breeding for sesame (Zhang et al. 2019; Zhang et al. 2021d). Five major steps are involved in (1) screening and creating elite breeding materials, (2) determining breeding objectives and hybrid crosses, (3) making breeding pipeline and selection strategy, (4) selecting elite progeny with highly efficient marker selection techniques, and (5) breeding superior elite sesame varieties. Thus, improving sesame varieties on nutrition-related traits should be performed according to the above strategy.

9.1 Marker-Assisted Breeding in Sesame

Genetic polymorphism analysis of nutrition-related traits among sesame accessions was initiated from molecular marker screening. In the past two decades, a variety of molecular markers, such as RAPD, AFLP, SCAR, SRAP, ISSR, SSR, EST-SSR, and SNP, have been developed for genetic diversity analysis, QTL mapping, and genetic dissection of key traits (Wei et al. 2013; Li et al. 2014; Wei et al. 2015; Yadav et al. 2022). The high variation explanation of these markers such as SSRs and SNPs presented the availability of marker-assisted breeding strategy. SSR markers have been demonstrated to play an important role in crossbreeding of major crops. While the SNP and InDel variants with the allelic nucleotide polymorphism of functional genes exhibit the genome variation character and can be more widely applied for sesame seed quality breeding (Murata et al. 2017; Zhou et al. 2022). Even though there are few reports on improving sesame varieties with high seed quality by the marker-assisted breeding approach now. Two SNP markers *SiDt27–1* and

SiSNPdwf1 for *SiDt* and *Sidwf1* gene, respectively, give the successful examples for new materials creation and molecular marker application in ideal variety breeding in sesame (Zhang et al. 2016; Miao et al. 2020).

9.2 Modern Breeding Technology of Sesame

In the past decade, HRSC, HAAS established the highly efficient germplasm creation and breeding platform for sesame, including EMS mutagenesis, distant hybridization, and genetic transformation. The mutant libraries, containing thousands of EMS mutants, transgenic lines, and interspecific progeny had been established. Of the above key techniques, genetic transformation was considered a cost-effective and powerful technology for the genetic improvement in sesame (Miao et al. 2021c). A great deal of effort has been devoted to establishing an efficient genetic transformation system for sesame. However, sesame is recalcitrant to regeneration in vitro (Bhaskaran and Jayabalan 2006). Application of genetic transformation technology lags behind. So many methods, including microparticle bombardment, and *Agrobacterium*-mediated and pollen tube channel methods, have been tried in the past decades. Except for pollen tube channel methods, the others depend on the tissue culture and shoot regeneration techniques (Miao et al. 2021c). Recent studies on tissue culture in sesame showed that adventitious shoots could be induced from the explants such as seed, shoot tip, stem, cotyledon, and young leaf via direct organogenesis way or indirect somatic embryogenesis way, which is dependent on the induction and differentiation of callus tissue (Miao et al. 2012; Miao et al. 2021c). Moreover, the regenerated sesame plants with suitable genotypes were induced from the adventitious buds producing in somatic embryogenesis way (Li et al. 1996; Miao et al. 2012). However, it is difficult to induce the adventitious shoots and the regenerated plants by callus induction and differentiation in sesame (Taskin and Turgut 1997). Two successful examples are that the regenerated plants respectively produce from adventitious shoots by direct explant induction for the hypocotyl and cotyledon of *S. radiatum* accession (Miao et al. 2012) and indirect callus differentiation for the de-embryonated cotyledon of an *S. indicum* variety TMV7 (Muthulakshmi et al. 2021).

Compared with other genetic transformation methods, *Agrobacterium*-mediated genetic transformation is more comprehensively studied in sesame. Taskin et al. (1999) discovered that sesame was susceptible to *A. tumefaciens*, but no transformed shoot was induced by *Agrobacterium*-mediated transformation in this study. Yadav et al. (2010) successfully induced cotyledon explants of sesame to produce transgenic plants carrying the neomycin phosphotransferase gene (*nptII*) and β-glucuronidase (*GUS*) gene for the first time, with a transformation frequency of 1.01%. Furthermore, Chowdhury et al. (2014) carried out a high-frequency *Agrobacterium*-mediated transformation by inducing the de-embryonated cotyledons at somatic embryogenic stages and presented the regeneration frequency of 57.33% and transformation efficiency of 42.66%. More recently, the transgenic sesame plants have been produced to introduce several functional genes related to key agronomic traits using the above method

(Muthulakshmi et al. 2021). Of these genes, acylCoA: diacylglycerol acyltransferase (*DGAT*) and phospholipid diacylglycerol acyltransferase (*PDAT*) play important roles in TAG biosynthesis, and *FAD3* are involved in polyunsaturated fatty acid biosynthesis. Overexpression of the *DGAT1* and *FAD3* and *PDAT1* and *FAD3* gene combinations showed to increase 9.0% and 11.5% TAG content in transgenic sesame leaves, respectively (Muthulakshmi et al. 2021). The study presents the modification potential of sesame plants as a biodiesel resource, even though whether the TAG content and fatty acid composition could be altered in the transgenic sesame seeds is still unclear. More attempts would be made to manipulate the functional genes by genetic transformation in sesame, so as to improve the nutrition-related traits in sesame. Predictably, with the abundance of functional gene resources, genetic transformation would show a huge function, especially in sesame nutrition improvement.

Recently, the metabolic engineering strategies were proposed to develop nutritionally optimized sesame oil by enhancing stearic acid (C18:0), oleic acid (C18:1), linolenic acid (C18:3), and total oil content (Bhunia et al. 2016). In the strategies, overexpression of acyl-ACP thioesterase gene or downregulation of stearoyl-acyl-carrier protein D9 FA desaturase (*SAD1*) gene were intended to enhance stearic acid (C18:0) content. Silence of *FAD2* would increase the oleic acid (C18:1) content. Moreover, overexpression of the heterologous *FAD3* and *FAD7* genes enhanced the accumulation of linolenic acid (C18:3) and manipulated the expression of the transcription factors Wrinkled 1 (*WRI1*), leafy cotyledon (*LEC1* and *LEC2*), and *DGAT* and *SiLTP1* genes and finally increased the content of total fatty acids in sesame. Definitely, more functional genes, especially transcription factors controlling the fatty acids and antioxidants metabolism, should be identified using the above techniques.

10 Conclusion and Future Prospects

In this chapter, we provided a summary of the progresses in sesame genetics and breeding research related to nutrition traits, as well as the achievement of key breeding materials and techniques. All the achievements reveal the importance of both the elite breeding materials and the molecular breeding elements for constructing the high-efficient molecular design breeding technology for sesame (Zhang et al. 2019; Zhang et al. 2021d). We also believe that sesame molecular breeding strategy would contribute to sesame seed quality breeding in the near future. Predictably, the superior sesame varieties introducing nutrition-related genes would be developed through the genetic engineering approach. Moreover, gene-editing system should be created for sesame based on the high-efficient transgene system.

Sesame is an ancient oilseed crop with high-quality nutrition, including abundant unsaturated fatty acids, proteins, tocopherol, and lignans. Of the main health-related components, the contents of oil, protein, fatty acid compositions, and lignans are hot research topics for sesame scientists around the world. In the past four decades, traditional breeding methods, including cross-breeding and mutagenesis breeding,

have achieved success in developing a series of sesame varieties with high yield and quality. With the development of the Sesame Genome Project and omics research, rich genome resources and tools have been obtained through continuous efforts. A list of molecular makers, QTLs, and functional genes related to key agronomic traits has been discovered in sesame and is available for molecular breeding programs (Zhang et al. 2021d). Even though the genetic dissection of oil and lignan contents and fatty acid compositions in sesame should be further developed, the designed molecular breeding strategies for sesame are going to play a key role in sesame genetics and breeding in the future. As regards the breeding techniques, the clustered regularly interspaced short palindromic repeats/associated protein 9 (CRISPR/Cas9) gene-editing system has been constructed for sesame with the *Agrobacterium*-mediated transformation recently (unpublished data, Zhang Haiyang). Differing from the transient expression validation of CRISPR/Cas9 vectors in hairy sesame root tissues reported by You et al. (2022), the stable transformation of CRISPR/Cas9 vectors proved the feasibility of gene edition and transgene system (Miao et al. 2021c) (unpublished data, Zhang Haiyang). In the future, the genetic modification approach will be employed to create mutants related to antinutritional genes and modify metabolic pathways of fatty acids and bioactive compounds to improve sesame varieties with low contents of free fatty acids and antinutritional factors such as phytic and oxalic acids, and high accumulation of antioxidant such as sesamin, sesamolin, and tocopherol (Langyan et al. 2022). In addition, the gene–metabolite relationships should be further investigated so as to uncover the regulation factors of the nutriment biosynthesis and accumulation in sesame. At present, only one metabolomics analysis was performed to reveal differences in metabolite accumulation between different sesame tissues (Dossou et al. 2021). Future exploration should develop comprehensive proteomics and metabolomics resources to understand fully the metabolic network of nutrient metabolites.

Acknowledgments This work was financially supported by the China Agriculture Research System of MOF and MARA (CARS-14), the Key Project of Science and Technology of Henan province (201300110600), the Key Research and Development Project of Henan Province (221111520400), and the Science and Technology Research Project of Henan Province (222102110081), the Zhongyuan Science and Technology Innovation Leading Talent Plan (214200510020), the Henan Province Specific Professor Position Program (SPPP2020), the Zhongyuan Scientist Workshop Construction (214400510026), and the Fund for Distinguished Young Scholars from Henan Academy of Agricultural Sciences (2022JQ01).

References

Andargie M, Vinas M, Rathgeb A, Möller E, Karlovsky P (2021) Lignans of sesame (*Sesamum indicum* L.): A comprehensive review. Molecules 26(4):883

Anilakumar KR, Pal A, Khanum F, Bawas AS (2010) Nutritional, medicinal and industrial uses of sesame (*Sesamum indicum* L.) seeds. Agric Conspec Sci 75:159–168

Bedigian D (2003) Evolution of sesame revisited, domestication, diversity and prospects. Genetic Res and Crop Evol 50(7):773–778

Bhaskaran P, Jayabalan N (2006) *In vitro* mass propagation and diverse callus orientation on *Sesamum indicum* L.—an important oil plant. J Agric Tech 2:259–269

Bhunia RK, Kaur R, Maiti MK (2016) Metabolic engineering of fatty acid biosynthetic pathway in sesame (*Sesamum indicum* L.): assembling tools to develop nutritionally desirable sesame seed oil. Phytochem Rev 15:799–811

Brigelius-Flohé R, Traber MG (1999) Vitamin E: function and metabolism. FASEB J 13(10): 1145–1155

Chowdhury S, Basu A, Kundu S (2014) A new high-frequency *Agrobacterium*-mediated transformation technique for *Sesamum indicum* L. using de-embryonated cotyledon as explant. Protoplasma 251:1–16

Cui C, Mei H, Liu Y, Zhang H, Zheng Y (2017) Genetic diversity, population structure, and linkage disequilibrium of an association-mapping panel revealed by genome-wide SNP markers in sesame. Front Plant Sci 8:1189

Dossa K, Diouf D, Wang L, Wei X, Zhang Y et al (2017) The emerging oilseed crop *Sesamum indicum* enters the "Omics" era. Front Plant Sci 8:1154

Dossou SSK, Xu F, Cui X, Sheng C, Zhou R et al (2021) Comparative metabolomics analysis of different sesame (*Sesamum indicum* L.) tissues reveals a tissue-specific accumulation of metabolites. BMC Plant Biol 21(1):352

Fukuda Y, Nagata M, Osawa T, Namiki M (1986) Contribution of lignan analogues to antioxidative activity of refined unroasted sesame seed oil. J Am Oil Chem Soc 63:1027–1031

Hashimoto M, Hossain S (2011) Neuroprotective and ameliorative actions of polyunsaturated fatty acids against neuronal diseases: beneficial effect of docosahexaenoic acid on cognitive decline in Alzheimer's disease. J Pharmacol Sci 116(2):150–162

He Q, Xu F, Min MH, Chu SH, Kim KW et al (2019) Genome-wide association study of vitamin E using genotyping by sequencing in sesame (*Sesamum indicum*). Genes Genom 41(9): 1085–1093

Herrera E, Barbas C (2001) Vitamin E: action, metabolism and perspectives. J Physiol Biochem 57 (2):43–56

Jiang Q, Elson-Schwab I, Courtemanche C, Ames BN (2000) gamma-tocopherol and its major metabolite, in contrast to alpha -tocopherol, inhibit cyclooxygenase activity in macrophages and epithelial cells. Proc Natl Acad Sci U S A 97(21):11494–11499

Jin UH, Lee JW, Chung YS, Lee JH, Yi YB et al (2001) Characterization and temporal expression of a ω-6 fatty acid desaturase cDNA from sesame (*Sesamum indicum* L.) seeds. Plant Sci 161:935–941

Joshi R, Kumar MS, Satyamoorthy K, Unnikrishnan MK, Mukherjee T (2005) Free radical reactions and antioxidant activities of sesamol: Pulse radiolytic and biochemical studies. J Agric Food Chem 53:2696–2703

Ju M, Miao H, Wang H, Zhang H (2021) Mutagenesis for Creation of Genetic Variability in Sesame. In: Miao H, Zhang H, Kole C (eds) The Sesame Genome. Compendium of Plant Genomes, Springer, Cham, pp 121–130

Kim MJ, Kim JK, Shin JS, Suh MC (2007) The *SeBHLH* transcription factor mediates transactivation of the *SeFAD2* gene promoter through binding to e- and g-box elements. Plant Mol Biol 64:453–466

Kim MJ, Go YS, Lee SB, Kim YS, Shin JS et al (2010) Seed-expressed casein kinase I acts as a positive regulator of the *SeFAD2* promoter via phosphorylation of the *SeBHLH* transcription factor. Plant Mol Biol 73(s4–5):425–437

Lahorkar P, Ramitha K, Bansal V, Narayana DA (2009) A comparative evaluation of medicated oils prepared using Ayurvedic and modified processes. Indian J Pharm Sci 71(6):656–662

Langyan S, Yadava P, Sharma S, Gupta NC, Bansal R et al (2022) Food and nutraceutical functions of sesame oil: An underutilized crop for nutritional and health benefits. Food Chem 389:132990

Li M, Zhang J, Wang Q, Yang YW, Wang ZS (1996) Callus induction and organogenesis of *Sesamum indicum* L. hypocotyl. J Henan Normal Uni 1:61–64

Li C, Miao HM, Wei LB, Zhang TD, Han XH et al (2014) Association mapping of seed oil and protein content in *Sesamum indicum* L. using SSR markers. PLoS ONE 9(8):e105757

Miao H, Ju M, Wei L, Ma Q, Zhang H (2012) Establishment of sesame callus induction and shoot regeneration system. Chin Bull Bot 47:162–170

Miao H, Li C, Duan Y, Wei L, Ju M et al (2020) Identification of a *Sidwf1* gene controlling short internode length trait in the sesame dwarf mutant dw607. Theor Appl Genet 133(1):73–86

Miao H, Sun Y, Wang L, Zhang H (2021a) Sesame Genome Assembly. In: Miao H, Zhang H, Kole C (eds) The Sesame Genome. Compendium of Plant Genomes, Springer, Cham, pp 225–238

Miao H, Sun Y, Li C, Wang L, Zhang H (2021b) Genome Annotation and Gene Families in Sesame. In: Miao H, Zhang H, Kole C (eds) The Sesame Genome. Compendium of Plant Genomes, Springer, Cham, pp 255–266

Miao H, Ju M, Wang H, Zhang H (2021c) Tissue Culture and Genetic Transformation in Sesame. In: Miao H, Zhang H, Kole C (eds) The Sesame Genome. Compendium of Plant Genomes, Springer, Cham, pp 131–144

Monteiro EM, Chibli LA, Yamamoto CH, Pereira MC, Vilela FM et al (2014) Antinociceptive and anti-inflammatory activities of the sesame oil and sesamin. Nutrients 6(5):1931–1944

Murata J, Ono E, Yoroizuka S, Toyonaga H, Shiraishi A et al (2017) Oxidative rearrangement of (+)-sesamin by CYP92B14 co-generates twin dietary lignans in sesame. Nat Commun 8(1): 2155

Murty DS (1975) Heterosis, combining ability and reciprocal effects for agronomic and chemical characters in sesame. Theor Appl Genet 45:294–299

Muthulakshmi C, Sivaranjani R, Selvi S (2021) Modification of sesame (*Sesamum indicum* L.) for triacylglycerol accumulation in plant biomass for biofuel applications. Biotechnol Rep (Amst) 32:e00668

Nakai M, Harada M, Nakahara K, Akimoto K, Shibata H et al (2003) Novel antioxidative metabolites in rat liver with ingested sesamin. J Agric Food Chem 51(6):1666–1670

Nimmakayala P, Perumal R, Mulpuri S, Reddy UK (2011) Sesamum. In: Kole C (ed) Wild crop relatives: genomic and breeding resources. Volume Oilseeds. Springer, Berlin Heidelberg, pp 261–273

Noguchi A, Fukui Y, Iuchi-Okada A, Kakutani S, Satake H et al (2008) Sequential glucosylation of a furofuran lignan, (+)-sesaminol, by *Sesamum indicum* UGT71A9 and UGT94D1 glucosyltransferases. Plant J 54(3):415–427

Ogata N, Kato M (2016) alf-diallel analysis for sesamin and sesamolin contents of sesame (*Sesamum indicum* L.) seeds. Japanese J Crop Sci 85:302–308

Ono E, Nakai M, Fukui Y, Tomimori N, Fukuchi-Mizutani M et al (2006) Formation of two methylenedioxy bridges by a Sesamum CYP81Q protein yielding a furofuran lignan (+)-sesamin. Proc Natl Acad Sci U S A 103(26):10116–10121

Ono E, Waki T, Oikawa D, Murata J, Shiraishi A et al (2020) Glycoside-specific glycosyltransferases catalyze regio-selective sequential glucosylations for a sesame lignan, sesaminol triglucoside. Plant J 101(5):1221–1233

Parker RS, Sontag TJ, Swanson JE (2000) Cytochrome P4503A-dependent metabolism of tocopherols and inhibition by sesamin. Biochem Biophys Res Commun 277:531–534

Pathak K, Rahman SW, Bhagawati S, Gogoi B (2017) Sesame (*Sesamum indicum* L.), an underexploited oil seed crop: Current status, features and importance - A review. Agric Rev 38(3): 223–227

Plietz P, Damaschun G, Zirwer D, Gast K, Schwenke KD et al (1986) Shape and quaternary structure of alpha-globulin from sesame (*Sesamum indicum* L.) seed as revealed by small angle x-ray scattering and quasi-elastic light scattering. J Biol Chem 261:12686–12691

Prakash K, Naik SN (2014) Bioactive constituents as a potential agent in sesame for functional and nutritional application. J Bioresour Eng Technol 1:48–66

Rangkadilok N, Pholphana N, Mahidol C, Wongyai W, Saengsooksree K et al (2010) Variation of sesamin, sesamolin and tocopherols in sesame (*Sesamum indicum* L.) seeds and oil products in Thailand. Food Chem 122(3):724–730

Ray RS, Katyal A (2016) Myeloperoxidase: Bridging the gap in neurodegeneration. Neurosci Biobehav Rev 68:611–620

Shenoy RR, Sudheendra AT, Nayak PG, Paul P, Kutty NG et al (2011) Normal and delayed wound healing is improved by sesamol, an active constituent of *Sesamum indicum* (L.) in albino rats. J Ethnopharmacol 133(2):608–612

Smith DE, Salerno JW (1992) Selective growth inhibition of a human malignant melanoma cell line by sesame oil *in vitro*. Prostaglandins Leukot Essent Fatty Acids 46(2):145–150

Song SN, You J, Shi LS, Sheng C, Zhou WY et al (2021) Genome-wide analysis of *nsLTP* gene family and identification of *SiLTPs* contributing to high oil accumulation in sesame (*Sesamum indicum* L.). Int J Mol Sci 22(10):5291

Sumathi P, Muralidharan V (2014) Gene effects and inheritance of branching and other yield attributing characteristics in sesame (*Sesamum indicum* L). Trop Agric Res Ext 16:92–101

Tashiro T, Fukuda Y, Osawa T, Namiki M (1990) Oil and minor components of sesame (*Sesamum indicum* L.) strains. J Am Oil Chem Soc 67(8):508–511

Taskin KM, Turgut K (1997) *In vitro* regeneration of sesame (*Sesamum indicum* L.). Turk J Bot 21: 15–18

Taskin KM, Ercan AG, Turgut K (1999) *Agrobacterium tumefaciens*-mediated transformation of sesame (*Sesamum indicum* L.). Turk J Bot 23:291–296

Teboul N, Gadri Y, Berkovich Z, Reifen R, Peleg Z (2020) Genetic architecture underpinning yield components and seed mineral-nutrients in sesame. Genes (Basel) 11(10):1221

Tzen JTC (2021) Beneficial Components in Sesame Proteins and Oil. In: Miao H, Zhang H, Kole C (eds) The Sesame Genome. Compendium of Plant Genomes, Springer, Cham, pp 59–78

Wan Y, Li HX, Fu GM, Chen XY, Chen F et al (2015) The relationship of antioxidant components and antioxidant activity of sesame seed oil. J Sci Food Agric 95(13):2571–2578

Wang LH, Zhang YX, Li PW, Zhang W, Zhang XF (2013) Variation of sesamin and sesamolin contents in sesame cultivars from China. Pak J Bot 45(1):177–182

Wang LH, Yu S, Tong CB, Zhao YZ, Liu Y et al (2014) Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. Genome Biol 15:R39

Wang L, Yu J, Zhang Y, You J, Zhang X et al (2021) Sinbase 2.0: An updated database to study multi-omics in *Sesamum indicum*. Plants (Basel) 10(2):272

Wang X, You J, Liu AL, Qi X, Li DH et al (2022) Variation in melatonin contents and genetic dissection of melatonin biosynthesis in sesame. Plants (Basel) 11(15):2005

Wei LB, Zhang HY, Zheng YZ, Guo WZ, Zhang TZ (2008) Developing EST-derived micro-satellites in sesame (*Sesamum indicum* L.). Acta Agron Sin 34(12):2077–2084

Wei WL, Zhang YX, Lü HX, Li DH, Wang LH et al (2013) Association analysis for quality traits in a diverse panel of Chinese sesame (*Sesamum indicum* L.) germplasm. J Integr Plant Biol 55(8): 745–758

Wei X, Liu KY, Zhang YX, Feng Q, Wang LH et al (2015) Genetic discovery for oil production and quality in sesame. Nat Commun 6:8609

Wei X, Zhu XD, Yu JY, Wang LH, Zhang YX et al (2016) Identification of sesame genomic variations from genome comparison of landrace and variety. Front Plant Sci 7:1169

Wei L, Miao H, Duan Y, Zhang H (2021) Classical Genetics of Sesame. In: Miao H, Zhang H, Kole C (eds) The Sesame Genome. Compendium of Plant Genomes, Springer, Cham, pp 79–120

Wu K, Yang MM, Liu HY, Tao Y, Mei J et al (2014) Genetic analysis and molecular characterization of Chinese sesame (*Sesamum indicum* L.) cultivars using insertion-deletion (InDel) and simple sequence repeat (SSR) markers. BMC Genet 15:35

Wu K, Wu WX, Yang MM, Liu HY, He GC et al (2017) QTL mapping for oil, protein and sesamin contents in seeds of white sesame. Acta Agron Sin 43(7):1003–1011

Xu FT, Zhou R, Cui XH, Gao Y, Zhang YX et al (2020) Variability and heterosis of sesamin and sesamolin contents in sesame (Sesamum indicum L.). Guangxi Agric Sci 51(10):2420–2428

Xu F, Zhou R, Dossou SSK, Song S, Wang L (2021) Fine mapping of a major pleiotropic QTL associated with sesamin and sesamolin variation in sesame (*Sesamum indicum* L.). Plants (Basel) 10(7):1343

Yadav M, Chaudhary D, Sainger M, Jaiwal PK (2010) Agrobacterium tumefaciens-mediated genetic transformation of sesame (*Sesamum indicum* L.). Plant Cell Tissue Organ Cult 103: 377–386

Yadav R, Kalia S, Rangan P, Pradheep K, Rao GP et al (2022) Current research trends and prospects for yield and quality improvement in sesame, an important oilseed crop. Front Plant Sci 13: 863521

Yamashita K, Nohara Y, Katayama K, Namiki M (1992) Sesame seed lignans and gamma-tocopherol act synergistically to produce vitamin E activity in rats. J Nutr 122(12):2440–2446

You J, Li D, Yang L, Dossou SSK, Zhou R et al (2022) CRISPR/Cas9-mediated efficient targeted mutagenesis in sesame (*Sesamum indicum* L.). Front. Plant Sci 13:935825

Yu J, Golicz AA, Lu K, Dossa K, Zhang Y et al (2019) Insight into the evolution and functional characteristics of the pan-genome assembly from sesame landraces and modern cultivars. Plant Biotechnol J 17(5):881–892

Yuan QL, Maio HM, Ma Q, Ju M, Zhang XT et al (2022) Establishment of near infrared spectral model and characteristic analysis of main components of fatty acids in sesame seeds. J Hebei Agric Sci 26(2):102–108

Yue W, Wei LB, Zhang TD, Li C, Miao HM et al (2012) Genetic diversity and population structure of germplasm resources in sesame (*Sesamum indicum* L.) by SSR markers. Acta Agron Sin 38 (12):2286–2296

Zhang YX, Zhang XR, Hua W, Wang LH, Che Z (2010) Analysis of genetic diversity among indigenous landraces from sesame (*Sesamum indicum* L.) core collection in China as revealed by SRAP and SSR maikers. Genes Genom 32:207–215

Zhang H, Wei L, Miao H, Zhang T, Wang C (2012) Development and validation of genic-SSR markers in sesame by RNA-seq. BMC Genomics 13:316

Zhang H, Miao H, Wang L, Qu L, Liu H et al (2013) Genome sequencing of the important oilseed crop *Sesamum indicum* L. Genome Biol 14(1):401–409

Zhang H, Miao H, Li C, Wei L, Duan Y et al (2016) Ultra-dense SNP genetic map construction and identification of *SiDt* gene controlling the determinate growth habit in *Sesamum indicum* L. Sci Rep 6:31556

Zhang H, Miao H, Ju M (2019) Potential for Adaptation to Climate Change Through Genomic Breeding in Sesame. In: Kole C (ed) Genomic Designing of Climate-smart Oilseed Crops. Springer, Cham, pp 374–376

Zhang H, Langham D, Miao H (2021a) Economic and Academic Importance of Sesame. In: Miao H, Zhang H, Kole C (eds) The Sesame Genome. Compendium of Plant Genomes, Springer, Cham, pp 1–18

Zhang H, Wang L, Miao H, Sun Y (2021b) Genome Sequencing of the Wild Sesamum Speices. In: Miao H, Zhang H, Kole C (eds) The Sesame Genome. Compendium of Plant Genomes, Springer, Cham, pp 255–266

Zhang H, Langham DR, Zhao Y, Khalafalla A, Miao H (2021c) Traditional Breeding in Sesame. In: Miao H, Zhang H, Kole C (eds) The Sesame Genome. Compendium of Plant Genomes, Springer, Cham, pp 145–158

Zhang H, Langham DR, Miao H (2021d) Prospect of Designed Breeding in Sesame in the Post-Genomics Era. In: Miao H, Zhang H, Kole C (eds) The Sesame Genome. Compendium of Plant Genomes, Springer, Cham, pp 291–296

Zhang RY, Gao JH, Shi YL, Lan YF, Liu HM et al (2022) Characterization of structure and antioxidant activity of polysaccharides from sesame seed hull. Front Nutr 9:928972

Zhou WY, Song SN, Dossou SSK, Zhou R, Wei X et al (2022) Genome-wide association analysis and transcriptome reveal novel loci and a candidate regulatory gene of fatty acid biosynthesis in sesame (*Sesamum indicum* L.). Plant Physiol Biochem 186:220–231

Zuo Y, Peng C, Liang Y, Ma KY, Chan HY et al (2013) Sesamin extends the mean lifespan of fruit flies. Biogerontology 14:107–119

Nutraceutical Usages and Nutrigenomics of Castor

Jasminkumar Kheni and Rukam S. Tomar

Contents

Abstract

Castor, *Ricinus communis* L., is one of the oldest and most significant non-edible oilseed crops. It is an important member of the Euphorbiaceae family of medicinal plants. The present chapter evaluates phytochemistry, biological and

J. Kheni · R. S. Tomar (✉)
Department of Biotechnology, Junagadh Agricultural University, Junagadh, Gujarat, India
e-mail: rukam@jau.in

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_18

pharmacological activity, and ethnomedicinal usage of castor. Castor contains various chemical components, which includes flavonoids, phenolic compounds, fatty acids, amino acids, terpenoids, phytosterol, etc. According to studies, these substances have anticonception, antidiabetic, antifertility, anti-inflammatory, antibacterial, antioxidant, hepatoprotective, insecticidal, and wound-healing properties. Additionally, the castor has demonstrated Hg and free radical scavenging activities as well as repellent qualities. Due to their use as medicines, castor plants have gained significant importance. The stem can be used to treat cancer and hypoglycemia, while the leaves can treat biliousness, antiviral, burns, ear and headache problems, night blindness, and malaria. You can apply the flowers to ease glandular and vaginal pain. In addition to cancer, piles, liver, and spleen diseases are all treated with fruits. In the treatment of rheumatoid arthritis, ascites, leprosy, rectal aches, hypoglycemia, asthma, and bronchitis, root can be used as a purgative. The castor oil plant has also shown negative effects since it contains the proteins ricin and ricinine which are highly poisonous. However, there is still a need for additional study to be done in relation to its medicinal significance and the active chemicals responsible for a variety of functions.

Keywords

Castor · *Ricinus communis* · Nutraceutical properties · Pharmacological properties

1 Introduction

Castor (*Ricinus communis* L.) belongs to the spurge (Euphorbiaceae) family. It is believed that it came from India and Ethiopia. Ricinus is a single species and *R. communis* is the only species. *Ricinus* and *Communis* are Latin words; "ricinus" is a tick, a specific synonym of the Mediterranean sheep tick (*Ixodes ricinus*), and is the common name for this tick's common species. The name Ricinus was invented by Caroleus Linnaeus, father of the taxonomy, because the seeds of the castor bean resemble "ticks" that inflict blood (a body that is mutilated by certain ticks), especially large ticks that are expanded with blood. Although plants and seeds are often referred to as Castor beans, they are not legumes. Castor is a member of the monotypic subtribe Ricininae of the genus Ricinus. It is one of the oldest plants and is becoming increasingly important as a crop for tropical and subtropical nations worldwide. Castor is a robust plant that yields 350–900 kg of oil per hectare, needs little maintenance, tolerates marginal soils, is simple to start in the field, and is drought-resistant. Castor oil exhibits excellent functional value in the pharmaceutical, industrial, and energy sectors. It has seen an increase in demand in recent years on the global market for its more than 700 applications, which range from biodiesel, plastic, and lubricants to medicine and cosmetics. Due to its high and low temperature tolerance capabilities, the oil is important for many industrial uses in comparison to other oils derived from plants. The majority of castor is grown in dry and semi-arid climates. It is grown

commercially in 30 nations, with India, China, Brazil, Russia, Thailand, Ethiopia, and the Philippines being the top producers of castor seeds, accounting for over 88% of global output. Castor seeds are extremely valuable and contain between 48% and 52% of the oil. Other plant components, including leaves, branches, flowers, fruits, and roots, can be utilized separately or in conjunction with other items for a variety of medical applications. Since at least 5000 years, India has used numerous plants in its traditional medical practices (Ayurvedic, Unani, and Siddha). More than 8000 herbal medicines have been documented in Ayurveda. Indian traditional medicines, folk medicines, and herbal remedies have utilized almost 6000 plants (Huxley 1984) for the treatment of different diseases. In most medical treatments, various parts of castor plants and the oil have been used as the base material. For example, leaves can be used to treat antivirals, bile ducts, burns, ear and head pain, malaria, and night blindness, stems are used to cure cancer and hypoglycemia. Flowers are effective for treating glandular and vaginal pain. Malignant diseases, liver and spleen ailments, and piles are all treated with fruits. Rheumatoid arthritis, ascites, asthma, bronchitis, diarrhea (gas from the stomach and intestines), hypoglycemia, and leprosy are all treated with the root bark as a cleanser (Neogi et al. 1989).

2 Nutraceutical Application of Castor Oil

The eyeball is covered and shielded by the conjunctiva, a delicate and thin membrane. An inflammation of the conjunctiva known as conjunctivitis is characterized by redness and typically ends with an ocular discharge of water. When eyes are exposed frequently or constantly to environmental factors and microorganisms that could cause infections or allergic reactions, it is a common eye ailment. Depending on the type of organism or agent involved and the intensity of the symptoms, it might be acute or chronic. Close physical contact is the most straightforward way to transfer conjunctivitis to others, especially among children (Pewitt 2004). The leaf decoction of *Achyranthes aspire*, mixed with castor oil, can be applied to the head and body 1 hour before bathing to overcome the problem of conjunctivitis.

A type of chronic skin inflammation characterized by weeping vesicular lesions, redness, and itching is called eczema. Other symptoms of eczema include skin edema/swelling, dryness, itchiness, crusting, peeling, blistering, breaking, leaking, or bleeding (Johannes et al. 2006; Williamson 2002). To treat eczema, castor oil is combined with powdered Indian birthwort (*Aristolochia indica*), and the oil is made by boiling *Datura stramonium* leaf juice, which is then applied to the skin.

When combined with copper sulfate, castor oil is used to treat a variety of skin conditions. Castor oil is combined with crushed *Alangium salvifolium* leaves, which are then applied to the area of inflammation. The scalp, ears, genitalia, and skin are the most typical areas of the body affected by the chronic skin disorder psoriasis. It is characterized by dry, red areas that are coated in scales. Cow urine in combination with the rhizome of *Curcuma domestica*, the seed of *Piper nigrum*, and the leaf of *Aristalochia bracteata* is combined to make a paste that is then fried in castor oil. The area with psoriasis should receive regular applications of this mixture for its treatment.

Sexually transmitted diseases (STDs) are also known as sexually transmitted infections (STIs) and are transmitted by human sexual behavior. Some are transmitted through childbirth or breastfeeding and through the re-use of medical needles which have been used by the infected people. STDs reduce the white and red blood cell counts (WBC and RBC), which leads to a weaker immune system. Castor oil packs improve WBC and RBC counts within 2 weeks, which strengthens the immune system of the body. To treat syphilis and gonorrhea, castor oil is used along with arsenic and copper sulfate. Although several drugs can increase lymph flow, castor oil can be applied topically to achieve the same goal. The number of lymphocytes in the blood increases when castor oil is absorbed by the skin. This results in a positive effect on the thyroid gland and/or lymphatic tissues. Increased lymph flow throughout the body accelerates toxins' removal, reduces swelling lymph node size, and improves overall organ function.

Castor oil is a widely used drug used to treat painful diarrhea and is often referred for constipation. In rural areas, children with diarrhea receive a small amount of castor oil that promotes gastrointestinal movements.

3 Nutraceutical Application of Castor Leaf

The condition known as yellow fever is caused by the excess bilirubin in the body, causing yellow skin and eyes. When the liver decomposes dead red cells, yellow pigments called bilirubin are formed. Jaundice is a sign of liver, gall bladder, or pancreas dysfunction. Tender castor leaf paste and coconut water are administered orally to patients with jaundice.

A fine paste of mature leaves is prepared to which a small amount of salt is added and then warmed up. This material is applied to the swollen muscles. In addition to calming the swelling, this also eases the pain, backache, and constipation.

Sesame oil is applied to the entire castor leaf and warmed on the metal plate. This is used on the gouty arthritis-affected joints. If the process is repeated every day for a week, it reduces pain and edema.

4 Nutraceutical Application of Castor Seed

Filariasis is a tropical parasitic and infectious disease caused by filariasis nematode worms and transmitted by the bite of a mosquito. Elephant edema with skin and subcutaneous tissue thickening is the most striking symptom. It affects mainly the lower extremities, with less frequent cases of the ear, mucous membranes, and amputation of skin. To treat filariasis, castor seed paste is given to the affected area.

5 Nutraceutical Application of Castor Root

Roots are used to treat swelling, fever, abdominal disease, arthritis, rheumatoid arthritis, pain in the back region, and similar diseases. A decoction of the dried roots is created using 20–25 grams. Alternatively, medicinal milk can also be made

from the roots. A dose of 40 ml of the decoction is given twice daily. This aids in easing the discomfort brought on by conditions like sciatica and backaches, among others. Additionally, it also soothes constipation.

6 Role of Castor in Cancer Research

The medicinal potential of ricin is considerable, and it is even being researched in the context of nanoparticle formulations for tumor therapy as an anticancer medication, in bone marrow transplantation, or in cell-based research. Based on the description of ricin's chemical structure and mode of action, the cytotoxic effects of ricin on tumor cells have already been seen in numerous investigations.

Lin and coworkers expanded on the first findings in rat sarcomas by describing the longer life of mice with Ehrlich ascites tumors after ricin treatment (Lin et al. 1973), as well as in a mouse model using human xenografts (FODSTAD and OLSNES 1977). Cancer cells have been found to be more sensitive to ricin, which may be due to their higher rate of protein synthesis, which makes them more susceptible to ricin's ability to inhibit this process, or to the presence of more ricin (B-chain) receptors on the surface of tumor cells, which facilitates better uptake. A preliminary phase of research in 54 patients with advanced cancer showed some therapeutic results in a small number of individuals (Fodstad et al. 1984). However, it was shown that ricin binds to almost with all cell types, making this binding specificity insufficient (Audi et al. 2005). Additionally, it was discovered that the ricin B chain played a part in the membrane-bound A chain's intracellular movement and translocation as well as its own pro-apoptotic actions (Timar et al. 1991). As a result, it was discovered that immunotoxins based on the conjugation of the entire ricin molecule, as opposed to only its A chain, were more effective (Olsnes 2004) Aside from the castor applications listed in the preceding paragraphs, Table 1 and Fig. 1 list a few of the other pharmacological features of the castor plant.

7 Nutraceutical Improvement in Castor: Molecular Basis

Many nations consider castor beans to be a significant crop. Castor oil, which is widely used in many industries, is the principal usage for it. After the oil is extracted, meals and cakes high in protein are what are left. They show promise for usage as protein supplements in the creation of feed. The presence of poisons, however, places restrictions on it. These are ricinine alkaloid and ricin protein. Genomic editing with the CRISPR/Cas9 system is one such technique. The technique entails using a small guide RNA fragment to help the Cas9 enzyme cut the target region of intact DNA. The delivery of Cas9 and the guide RNA genes to the cells of the edited plant is often carried out by plasmid vectors. Promoting sequences with intact USE and TATA boxmotifs have been used in the construction of CRISPR/Cas9 vectors that can be used to effectively edit the castor bean genes involved in the synthesis of ricin and ricinine (Alexandrov and Karlov 2021).

Table 1 Pharmacological properties of different parts of castor

S. No	Pharmacological activities	Part used	Mode/Reaction	Reference
1	Abortifacient effect	Castor oil	Anti-implantation and anti-ovulation effects due to ricin-A	Salhab et al. (1999)
2	Analgesic activity	Root bark extract	Saponin, steroids, and alkaloids	Rajeshkumar et al. (2013)
3	Acid phosphatase stimulation	Castor oil	The intraluminal acid phosphatase increase	Ross (2003)
4	Anti-fertility activity	Seed extract	Anti-fertility effects due to steroids and alkaloids	Jena and Gupta (2012)
5	Antiasthmatic activity	Root extract	Saponin and flavonoids have a property to stabilize and relax smooth muscle	Jena and Gupta (2012)
6	Anticonceptive effect	Seed extract	Direct effect on the uterus	Okwusasaba et al. (1997)
7	Anti-amoebic activity	Root/stem extract		Ross (2003)
8	Antimicrobial activity	Root extract	Activity against pathogenic fungi	Jena and Gupta (2012)
9	Antidiabetic activity	Root extract	Antihyperglycemic activity in one of the root fractions	Shokeen et al. (2008)
10	Antibacterial activity	Leaf extracts	Activity against pathogenic bacteria	Islam et al. (2010), Jombo and Enenebeaku (2007)
11	Anticancer effects	Seed and stem extract	In vitro anticancer effects against cancer cell lines: Colon cancer cell line, liver cancer cell line, breast cancer cell line, cervix cancer cell line	Prakash and Gupta (2014)
12	Anti-inflammatory activity	Root extract	Anti-inflammatory and free radical scavenging activities of the root extract of *R. communis* in Wistar albino rats	Nath et al. (2011)
13	Antioxidant activity	Root extract	The phytochemicals constituents responsible for antioxidant activity may be methyl ricinoleate, Ricinoleic acid, 12 octadecadienoic acid and methyl ester	Jena and Gupta (2012)
14	Central nervous system (CNS) stimulant	Seed coat extract	Ricinine is responsible for central nervous system stimulant	Williamson (2002)

Genetic markers including agromorphological traits, biochemical markers, and cytological markers were commonly utilized to characterize genetic variation in the germplasm from India, Nigeria, Turkey, China, Brazil, Iran, and Ethiopia. Depending on the markers and genotypes examined, these markers suggested a low-to-high

Fig. 1 Pharmacological usage of different parts of castor

amount of variability in the castor bean germplasm. Castor bean genetic diversity was assessed using both dominant and codominant molecular markers, such as expressed sequence tag-simple sequence repeats (EST-SSR), start codon targeted, inter-simple sequence repeats, random microsatellite amplified polymorphic DNA (RMAPD), amplified fragment length polymorphism (AFLP), and simple sequence repeats (SSR). Utilizing molecular markers to characterize the genetic diversity in germplasm from countries like India, China, Brazil, Mexico, and the worldwide collection, castor bean exhibits a low-moderate level of genetic variability. The availability of the castor bean genome has accelerated the creation of straightforward, rapid, and trustworthy DNA markers for the evaluation of genetic variation in the plant. Additionally, these markers will be an important tool for determining genetic diversity and supporting projects for marker-assisted breeding to improve crops.

Castor organelle genomics revealed that single nucleotide polymorphism markers and phylogenetic analysis allowed the identification of two significant clades that were not visible in earlier population genetic research utilizing genetic markers obtained from nuclear DNA. Within each major clade, two separate subclades could be identified, and extensive genotyping of castor bean populations around the world verified the previously noted low levels of genetic variation and revealed that each sub-clade was widely distributed geographically (Rivarola et al. 2011).

It has been determined that the accessions from East Africa are the extant wild progenitors of castor bean and that domestication took place around 3200 years ago using a de novo chromosome-level genome assembly approach of the wild progenitor of castor bean with resequencing and analyzing 505 worldwide accessions. This study

also demonstrated a link between the Turkana depression's previous temperature variability and the large genetic divergence between wild populations in Ethiopia and Kenya. The present work, in conjunction with quantitative trait locus analysis, has discovered significant candidate genes linked to plant architecture and seed size (Xu et al. 2021). Whereas the castor bean's draft genome sequencing revealed critical genes for significant oil synthesis, unexpectedly, there are more members of the ricin gene family. According to comparative genomics research, the dicotyledonous lineage has been preserved through an ancient hexaploidization event (Chan et al. 2010).

8 Other Uses of Castor

8.1 Acid Phosphatase Stimulation

Rats were administered 2 ml/animal of castor seed oil intragastrically, which increased intraluminal acid phosphatase liberation in the duodenum and jejunum but not in the stomach (Ross 2003).

8.2 Analgesic Activity

While the common medication diclofenac was utilized at a dose of 50 mg/kg for analgesic action, the aqueous extract of the root bark of *R. communis* was examined at doses of 100 mg/kg and 200 mg/kg. Six albino mice of each sex were used in each group for the evaluation study using the hot plate method and tail immersion method. The findings showed that the extract had notable anti-nociceptive action against the two mouse pain models. Saponin, steroids, and other phytochemical substances may be responsible for the same.

8.3 Antidiabetic Activity

The anti-diabetic potential of *R. communis* root extract (RCRE) was investigated. The diabetic rats received RCRE (500 mg/kg b.w.) for 20 days in order to achieve this goal. The outcome showed beneficial effects on liver and kidney functions as well as the total lipid profile and fasting blood glucose. Of the total fractions studied, only one (R-18) showed a significant amount of anti-hyperglycemic activity. *R. communis* can therefore be a crucial component of an efficient phytomedicine against diabetes (Shokeen et al. 2008).

8.4 Antimicrobial Activity

The antibacterial properties of several solvent extracts of *R. communis* roots (200 mg/ml) were examined. In this regard, the pathogens such as *Aspergillus*

niger, *Bacillus subtilis*, *Candida albicans, E. coli*, *Proteus vulgaris*, *Pseudomonas aeruginosa*, *S. aureus*, and *Salmonella typhimurium* that cause different disorders were studied using the well-known diffusion method. Maximum antibacterial activity was demonstrated by the hexane and methanol extracts, but the aqueous extracts displayed reduced antimicrobial characteristics (Jena and Gupta 2012).

8.5 Anticancer Effects

The ethanolic extract of *R. communis* was tested against seven human cancer cell lines for its in vitro anticancer properties. Colon cancer cell lines (Colon HT-29, SW-20, SiHa), liver cancer cell lines (Hep-2), breast cancer cell lines (T-47D), cervical cancer cell lines (OVCAR-5), and prostate cancer cell lines were the seven cell lines (PC-3). The test substance was tested for cytotoxicity against all cell lines at a concentration of 100 g/ml using the sulforhodamine B (SRB) assay. The greatest activity was shown by the stem component of *Ricinus communis* against SiHa (47%), whereas the ethanolic extract of the seed showed 41% activity against Colon 502713 (Prakash and Gupta 2014).

8.6 Anti-inflammatory Activity

A test was done on Wistar albino rats to see how well *R. communis* root extract reduced inflammation and scavenged free radicals. The study found that the aforementioned extract was effective for the activities examined. The flavonoids, alkaloids, and tannins contained in the plant may be the chemical components that triggered the aforementioned behaviors (Nath et al. 2011).

8.7 Bone Regeneration Activity

The purpose of the study was to assess the biocompatibility and potential of the Ricinus communis polyurethane to stimulate bone regeneration. The findings showed that RCP is a biocompatible substance that can promote matrix mineralization when mixed with calcium carbonate or calcium phosphate. Alkaline phosphatase addition and subsequent incubation of RCP in synthetic bodily fluid may enhance its biological characteristics. RCP has the benefit over demineralized bone in that the latter's reabsorption process is slower (Rana et al. 2012).

8.8 Repellent Properties

Castor oil has been used as a spray to deter moles, despite its unpleasant taste. Castor plants have also been utilized as a repellant by being grown around areas where

moles are a problem because of their foul odor. Care should be taken that children and pets have no access to the poisonous seeds (Baker and Grant 2018).

9 Important Composition of Castor Oilseed

9.1 Phospholipids

Phospholipids are one type of lipid that creates lipid bilayers in the cell membrane. Tocopherols and phospholipids may work together to postpone the onset of lipid oxidation (Chew and Nyam 2019). The pharmaceutical and food sectors can benefit from knowing more about phospholipids, especially when used as emulsifiers.

Castor oilseed's phospholipid makeup has long been researched. Castor has less than 1% total polar lipids, with phosphatidylcholine, phosphatidylinositol, and phosphatidylethanolamine being the main constituents, according to Moreau et al. (Moreau et al. 1980). Phosphatidylcholine, which made up 30% of them, was the most common. Donaldson also noted that 2-day-old castor oilseed contained 2% phospholipid, and that the values for the various phospholipid classes were comparable (Donaldson 1976). The phospholipid values recorded for castor are low when compared to other oilseed crops, such as chufa nuts, which have a 5.4% phospholipid content (Oderinde and Tairu 1992). Dark-colored oils typically have high phospholipid concentrations as a result. Because of the low concentration of phospholipids in the oils, castor oil has a clear, pale yellow tint.

9.2 Phenolic Components

Secondary metabolites from several categories are present in phenolic substances. This substance has one or more hydroxyl groups that can attach to aromatic molecules directly (Rampadarath et al. 2014). The three main categories of dietary phenolic substances are tannins, phenolic acids, and flavonoids. The two principal subclasses of phenolic acids are hydroxycinnamic acid and hydroxybenzoic acid (Chew and Nyam 2019). Gallic, syringic, and vanillic acids are byproducts of hydroxybenzoic acids, which primarily have aromatic rings with C1-C6 structure (Boualem et al. 2017). Having regard to hydroxycinnamic acids, their derivatives with a C3-C6 structure include caffeic, ferulic, and p-coumaric acids. Varied oils have different phenolic component types and contents. The flavor and antioxidant properties of oils are influenced by phenolic substances.

9.3 Fatty Acids

Vegetable oils contain linolenic and oleic acids, which are very good for human health and have been used to cure a number of ailments like diabetes, skin cancer,

renal illness, heart attacks, lupus, high blood pressure, and high cholesterol levels (Ganesan et al. 2018). Because it resists oxidation, oleic acid can be employed to enhance the activities of antioxidants and as agents that prevent polymerization (Anjani 2012). It can be blended with other oils to stop oxidation when paired with antioxidants (such as tocopherols). Because they have certain beneficial effects for the skin, linoleic, palmitic, and stearic acids are utilized widely in the beauty industry (Ganesan et al. 2018). Cosmetic items or pharmaceuticals are made using stearic acid esters such as glycol distearate, glycol stearate, and ethylene glycol. Glycol distearate, glycol stearate, and ethylene glycol are stearic acid esters that are used to make cosmetic items or to enhance the pearly look of shampoos (Gunstone 2009). The presence of these fatty acids in castor oil suggests that this plant has nutritional and industrial benefits.

9.4 Insecticidal and Pesticidal Activity of Castor

Castor oil was compared against cottonseed, linseed, rosin, and petroleum spray oils for the treatment of many citrus pests, particularly red scale (*Chrysomphalus aurantii*) with a deadly immersion duration of 1400 min. Out of all the vegetable oils, castor oil outperformed the light petroleum distillate in terms of managing the scale. However, the vegetable oils were also more phytotoxic (De Ong et al. 1927).

Castor oil prevented the hatching of sweet potato whitefly eggs better than cottonseed, peanut, soybean, and sunflower oils, with only 19% of the eggs surviving (Fenigstein et al. 2001). The outcome was comparable to peanut oil and much superior to cottonseed, soybean, and sunflower oils. The same study showed that a 3% solution of castor oil was successful in lowering the survival rates of larvae, with survival rates for the first instar being 7.9%, the second instar being 4.1%, the third instar being 14.0%, and the fourth instar being 19.0%. A 3% mortality rate after 2 hours, but a 75% mortality rate after 24 hours, indicated that 3% castor oil was only slightly effective against humans.

Another research suggested that *B. tabaci* larvae fed with castor oil had marginally greater survival rates than those treated with cottonseed oil; however there was no statistically significant difference between the two treatments (Butler Jr et al. 1991). The stakes were made of wood from the *Acacia nilotica* tree and were treated with castor oil using vacuum pressure, surface application, and dipping at concentrations of 10%, 15%, and 20% using each of the three techniques. After being sun- and oven-dried, the stakes were buried in a pit with *Odontotermes obesus* termites. For all three application types, all rates, and all types of drying techniques, castor oil significantly reduced wood loss from termite damage compared to the untreated control (Ahmed et al. 2014). The 20% rate offered marginally but significantly greater control than the lower rates. The 20% concentration by the dipping process resulted in the minimum percent loss, with 8.91% loss in sap wood and 7.62% loss in heart wood. Losses in the control were over 40% for heart wood and over 50% for sap wood.

9.4.1 Improvement of Pharmacological Quality of Castor Oil Through Biotechnological Approaches

Plant biotechnologists have been interested in the metabolism of castor, since it can accumulate significant amounts of ricinoleic acid. The primary extra plastidial alteration of fatty acids in castor oil plants is the hydroxylation of oleate, which is catalyzed by the membrane-bound enzyme FAH12 hydroxylase, which is closely linked to FAD2 and FAD3 (Van De Loo et al. 1995). This enzyme can add a hydroxyl group to oleate at the delta 12 position, which is where the endosperm of castor oil seeds develops. Nevertheless, the activity of this enzyme by itself is unable to account for the fatty acid content of castor oil (Fig. 2).

Various experimental studies have shown that the castor seed's synthetic machinery is specifically designed to improve ricinoleic acid into TAGs, involving the entire route rather than just specific enzymes (Hu et al. 2012). The castor plant is a very interesting manufacturing platform for uncommon fatty acids, particularly hydroxylated fatty acids of economic significance, because of this biochemical

Fig. 2 Different techniques to improve castor neutraceuticals

characteristic. The castor plant, which can be grown as an annual crop and is feasible with both productive and low-maintenance, is of particular scientific interest (Baldwin and Cossar 2009).

Significant strides in the use of castor oil plants as biofactories were accomplished once the toxicity of the seeds was removed (Sousa et al. 2017). Additionally, a draft version of the plant genome has been released, exposing the genetic makeup of the plant (Chan et al. 2010). The synthesis of other specialized fatty acids in this plant may one day be based on castor oil mutants with changed fatty acid compositions, such as the OLE1 mutant with up to 80% oleic acid and decreased ricinoleic acid content (Venegas-Calerón et al. 2016).

A methodology based on the transformation of seed dissected embryos, followed by selection and plant regeneration, has also been used to produce transgenic castor oil plants (Ahn et al. 2007; Sujatha and Sailaja 2005). However, very little transformation efficiency (0.04%) was seen using meristem-based methods in this plant. Transient gene expression studies are frequently used to enhance plant physiology, biochemistry, and biotechnology research of some plants because permanent transformation of those plants is typically a difficult process with numerous drawbacks.

10 Conclusion

For at least 5000 years, numerous plants have been used in traditional medicine in India. In traditional Indian medicine, folk medicine, and herbal medicine, nearly 6000 plants have been used to treat various diseases. In most of the medical treatments, different parts of castor plants and oil were used as raw materials. The leaf decoction of *Achyranthes* aspire mixed with castor oil can be applied to the head and body one hour before bathing to solve the problem of conjunctivitis. For the treatment of eczema, castor oil can be applied to the skin in combination with powdered Indian birthwort. In combination with copper sulfate, castor oil is used to treat a variety of skin diseases. Castor oil is combined with crushed leaves of *Alangium salvifolium*, which are then applied to inflamed areas. Castor oil packs improve the number of white and red blood cells within two weeks, which strengthens the body's immune system. The number of lymphocytes in the blood increases when castor oil is absorbed by the skin. In rural areas, children with diarrhea are given a small amount of castor oil, which promotes gastrointestinal movements. Tender castor leaf paste and coconut water are given orally to patients with jaundice. This not only has a soothing effect on swelling but also relieves pain, back pain, and constipation. The medical potential of ricin is considerable, and it is even being explored in nanoparticle formulations for tumor therapy as an anticancer drug, in bone marrow transplantation, or in cell-based research. Based on the description of the chemical structure and mode of action of ricin, numerous studies have already established the cytotoxic effect of ricin on tumor cells. Immunotoxins based on conjugation of the entire ricin molecule, not just the A-chain, were found to be more effective. Because of their use as medicines, castor plants have become very important. However, more studies are needed on their medicinal importance and the active compounds responsible for a variety of functions.

References

Ahmed S, Fatima R, Nisar M, Hassan B (2014) Evaluation of castor bean oil on *Acacia nilotica* as wood preservative against *Odontotermes obesus* (Ramb.)(Termitidae: Isoptera). Int Wood Prod J 5:5–10

Ahn Y-J, Vang L, McKeon TA, Chen GQ (2007) High-frequency plant regeneration through adventitious shoot formation in castor (*Ricinus communis* L.). In Vitro Cell Dev Biol Plant 43:9–15

Alexandrov O, Karlov G (2021) Sequencing of the U6 promoters in castor beans and vector construction for CRISPR/Cas9 genomic editing on their basis. Sel'skokhozyaistvennaya Biol [Agric Biol] 56:20–31

Anjani K (2012) Castor genetic resources: a primary gene pool for exploitation. Ind Crop Prod 35: 1–14

Audi J, Belson M, Patel M, Schier J, Osterloh J (2005) Ricin poisoning: a comprehensive review. JAMA 294:2342–2351

Baker B, Grant J (2018) Castor oil profile – active ingredient eligible for minimum risk pesticide use. New York State Integrated Pest Management. Cornell University, Geneva, pp 1–10

Baldwin BS, Cossar RD (2009) Castor yield in response to planting date at four locations in the South-Central United States. Ind Crop Prod 29:316–319

Boualem M, Mokhtar M, Saiah F, Benourad F, Bouhadiba R, Berkani A (2017) Identification of *Mentha piperita* L. and *Ricinus communis* L. polyphenols by HPLC-DAD-ESI-MS and evaluation of their insecticidal properties against *Aphis spiraecola* P. South Asian J Exp Biol 7:28–34

Butler G Jr, Puri S, Henneberry T (1991) Plant-derived oil and detergent solutions as control agents for *Bemisia tabaci* and *Aphis gossypii* on cotton. SW Entomol 14:287–293

Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, Melake-Berhan A, Jones KM, Redman J, Chen G (2010) Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotechnol 28:951–956

Chew SC, Nyam KL (2019) Kenaf (*Hibiscus cannabinus* L.) seed oil. In: Ramadan MF (ed) Fruit oils: chemistry and functionality. Springer, Cham, pp 451–494

De Ong ER, Knight H, Chamberlin JC (1927) A preliminary study of petroleum oil as an insecticide for citrus trees. Hilgardia 2:251–384

Donaldson RP (1976) Membrane lipid metabolism in germinating castor bean endosperm. Plant Physiol 57:510–515

Fenigstein A, Eliyahu M, Gan-Mor S, Veierov D (2001) Effects of five vegetable oils on the sweetpotato whitefly *Bemisia tabaci*. Phytoparasitica 29:197–206

Fodstad O, Olsnes S (1977) Studies on the accessability of ribosomes to inactivation by the toxic lectins abrin and ricin. Eur J Biochem 74:209–215

Fodstad O, Kvalheim G, Godal A, Lotsberg J, Aamdal S, Host H, Pihl A (1984) Phase I study of the plant protein ricin. Cancer Res 44:862–865

Ganesan K, Sukalingam K, Xu B (2018) Impact of consumption and cooking manners of vegetable oils on cardiovascular diseases – a critical review. Trends Food Sci Technol 71:132–154

Gunstone F (2009) The chemistry of oils and fats: sources, composition, properties and uses. Blackwell Publishing Ltd/CRC Press, Boca Raton, pp 257–280

Hu Z, Ren Z, Lu C (2012) The phosphatidylcholine diacylglycerol cholinephosphotransferase is required for efficient hydroxy fatty acid accumulation in transgenic Arabidopsis. Plant Physiol 158:1944–1954

Huxley A (1984) Green inheritance: the world wildlife fund book of India. Collins/Harvel, London

Islam T, Bakshi H, Sam S, Sharma E, Hameed B, Rathore B, Gupta A, Ahirwar S, Sharma M (2010) Assessment of antibacterial potential of leaves of *Ricinus communis* against pathogenic and dermatophytic bacteria. Int J Pharm Res Dev 1:1–7

Jena J, Gupta AK (2012) *Ricinus communis* Linn: a phytopharmacological review. Int J Pharm Pharmaceut Sci 4:25–29

Johannes R, Bernhard P, Thomas R (2006) Handbook of atopic eczema. Springer-Birkhäuser Verlag, Berlin, p 4

Jombo G, Enenebeaku M (2007) Antimicrobial susceptibility patterns of bacteria to seed extracts of *Ricinus communis*: findings of a preliminary study in Nigeria. Intern J Microbiol 4(1):22–27

Lin JY, Chang YC, Huang LY, Tung TC (1973) The cytotoxic effects of abrin and ricin on *Ehrlich ascite*s tumor cells. Toxicon 11:379–380

Moreau RA, Liu KD, Huang AH (1980) Spherosomes of castor bean endosperm: membrane components, formation, and degradation. Plant Physiol 65:1176–1180

Nath S, Choudhury MD, Roychoudhury S, Talukdar AD, Sirotkin AV, Baková Z, Kádasi A, Maruniaková N, Kolesárová A (2011) Restorative aspect of castor plant on mammalian physiology: a review. J Microbiol Biotechnol Food Sci 1:236–246

Neogi B, Prasad M, Rao R (1989) Ethnobotany of some weeds of Khasi and Garo hills, Meghalaya, Northeastern India. Econ Bot 43:471–479

Oderinde R, Tairu A (1992) Determination of the triglyceride, phospholipid and unsaponifiable fractions of yellow nutsedge tuber oil. Food Chem 45:279–282

Okwusasaba F, Das S, Isichei C, Ekwenchi M, Onoryvwe O, Olayinka A, Uguru V, Dafur S, Ekwere E, Parry O (1997) The anticonceptive and the effect on uterus of ether extract, 18312-J of *Ricinus communis*. Phytother Res 10:97–100

Olsnes S (2004) The history of ricin, abrin and related toxins. Toxicon 44:361–370

Pewitt D (2004) Keep an eye toward the nose: these treatments can help stop the charge of rhinoconjunctivitis. Rev Optomet 141(6):125–127

Prakash E, Gupta D (2014) In vitro study of extracts of *Ricinus communis* Linn on human cancer cell lines. J Med Sci Pub Health 2(1):15–20

Rajeshkumar D, Nagachaitanya V, Manasa G, Usharani A, Nagaraju K (2013) Pharmacological evaluation of analgisic activity of aqueous extract of *Ricinus communis* root bark. Int J Toxicol Pharmacol Res 5(4):94–95

Rampadarath S, Puchooa D, Ranghoo-Sanmukhiya VM (2014) A comparison of polyphenolic content, antioxidant activity and insecticidal properties of *Jatropha* species and wild *Ricinus communis* L. found in Mauritius. Asian Pacific J Trop Med 7:384–S390

Rana M, Dhamija H, Prashar B, Sharma S (2012) *Ricinus communis* L.— a review. Int J Pharm Tech Res 4:1706–1711

Rivarola M, Foster JT, Chan AP, Williams AL, Rice DW, Liu X et al (2011) Castor bean organelle genome sequencing and worldwide genetic diversity analysis. PLoS One 6(7):e21743

Ross IA (2003) Medicinal plants of the world: volume 1 chemical constituents, traditional and modern medicinal uses. Humana Press Inc, New Jersey, pp 277–288

Salhab A, Shomaf M, Gharaibeh M, Amer N (1999) Effects of castor bean extract and ricin A-chain on ovulation and implantation in rabbits. Contraception 59:395–400

Shokeen P, Anand P, Murali YK, Tandon V (2008) Antidiabetic activity of 50% ethanolic extract of *Ricinus communis* and its purified fractions. Food Chem Toxicol 46(11):3458–3466

Sousa NL, Cabral GB, Vieira PM, Baldoni AB, Aragão FJ (2017) Bio-detoxification of ricin in castor bean (*Ricinus communis* L.) seeds. Sci Rep 7:15385

Sujatha M, Sailaja M (2005) Stable genetic transformation of castor (*Ricinus communis* L.) via *Agrobacterium tumefaciens*-mediated gene transfer using embryo axes from mature seeds. Plant Cell Rep 23(12):803–810

Timar J, McIntosh D, Henry R, Cumber A, Parnell G, Davies A (1991) The effect of ricin B chain on the intracellular trafficking of an a chain immunotoxin. Br J Cancer 64(4):655–662

Van De Loo FJ, Broun P, Turner S, Somerville C (1995) An oleate 12-hydroxylase from *Ricinus communis* L. is a fatty acyl desaturase homolog. Proc Natl Acad Sci U S A 92(15):6743–6747

Venegas-Calerón M, Sánchez R, Salas JJ, Garcés R, Martínez-Force E (2016) Molecular and biochemical characterization of the OLE-1 high-oleic castor seed (*Ricinus communis* L.) mutant. Planta 244(1):245–258

Williamson EM (2002) Major herbs of Ayurveda. Churchill Livingstone, London

Xu W, Wu D, Yang T, Sun C, Wang Z, Han B, Wu S, Yu A, Chapman MA, Muraguri S (2021) Genomic insights into the origin, domestication and genetic basis of agronomic traits of castor bean. Genome Biol 22(113):1–27

Genetic Enhancement of Nutraceuticals in Linseed: Breeding and Molecular Strategies

C. Manimurugan, A. Zanwar, and M. Sujatha

Contents

Abstract

Linseed (or flax) is an important industrial oilseed crop mainly grown in temperate climates for its oil and fiber properties. Linseed fiber separated from stalks is a bast fiber which possesses high mechanical properties. The textile "linen" made from this fiber is popularly used in the textile industry. Seeds of linseed contain 33–47% of oil, with excellent drying property, and its oil is mainly used in the manufacturing of paints, varnishes, linoleum, oil cloths, and printing inks. In the recent past, linseed has gained attention and considered as "superfood" because it is one of the richest sources of omega-3 fatty acid or alpha-linolenic acid (ALA), a nutritionally important fatty acid, phytochemical compounds, vitamins, and minerals. Genetic

C. Manimurugan · M. Sujatha (✉)
ICAR-Indian Institute of Oilseeds Research, Hyderabad, India

A. Zanwar
Interactive Research School for Health Affairs, Bharati Vidyapeeth (Deemed to be University), Pune, India

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_19

enhancement in linseed through conventional and biotechnological tools was focused mainly on increased productivity, enhancing oil content and quality, increasing seed size, early maturity, lodging resistance, plant height, and resistance to major diseases. Limited efforts were made toward improvement of the nutraceutical properties of the crop concerning the omega-3 fatty acid content, oil quality modification, essential lignan (secoisolariciresinol diglucoside – SDG), and mucilage which have proven value of linseed as a functional food. Attempts were made to assess variation for these components among the linseed accessions and to study the genotype and environment interaction based on quantitative traits and molecular markers. A noteworthy achievement was made in the development of solin flax (edible oil) through mutagenesis with a point mutation in the *LuFAD3A* and *LuFAD3B* genes encoding microsomal desaturases and expanding the utility of linseed oil for edible purposes. During the past two decades, efforts of different research groups have led to the enrichment of genomic resources in terms of mapping populations, construction of linkage maps, development of molecular markers, and identification of quantitative trait loci (QTLs) and quantitative trait nucleotides (QTNs) for traits of agronomic importance. Genetic engineering studies in flax were encouraging and modifications with regard to SDG content, fiber quality, tolerance to herbicides, and resistance to *Fusarium* were successfully demonstrated. Use of nanoparticles of linseed oil in assessing antitumor activity and as a source of omega-3 in various food preparations and hydrogel derived from mucilage for drug delivery in skin care products is receiving special attention as they are found to be safe with guaranteed delivery and maximum benefit. Despite the promise the crop holds and the research advancements made in this crop, there is a long way to go to exploit the full potential of the crop for diverse nutraceutical and pharmaceutical uses.

Keywords

Linseed · Alpha linolenic acid · Lignan · Nutraceutical uses · Linseed breeding · Genetic engineering

1 Introduction

Linseed is one of the important self-pollinated oilseed crops, which is mainly grown for its oil and fiber properties. Linseed with a chromosome number of $2n = 30$ is belonging to the family Linaceae and the genus *Linum*. The name "Linum" is originated from the Celtic word lin or "thread," and the name "usitatissimum" is Latin for "most useful." There is a lot of disagreement regarding the origin of linseed. Vavilov (1935) proposed "Mediterranean region" as the origin of cultivated annual linseed. Linseed was introduced in North America by colonists and cultivated in a larger area mainly for its fiber. It was a major cultivated fiber crop in ancient Egypt, and linen was considered as a symbol of purity in Egypt. The drying property of linseed oil was extensively utilized in paint industries. Globally, linseed is

cultivated under 3.54 million ha with production of 3.37 million tons and average productivity of 951 kg/ha. The countries including Kazakhstan, Russia, Canada, China, India, the United States, Ethiopia, the United Kingdom, Argentina, Germany, Afghanistan, and France are the major linseed growers in the world. Kazakhstan with 31.4% contribution is the leading producer followed by Russian Federation (23.4%) and Canada (17.2%) (FAOSTAT 2020) (Table 1). During 2020, the top exporters of linseed were Canada, Russia, Kazakhstan, Belgium, and Poland, while the top importers were China, Belgium, Germany, the United States, and Poland. Top linseed-consuming countries are China, Belgium, the United States, Turkey, and Germany. The highest levels of linseed oil per capita consumption were registered in Belgium, followed by Turkey, Germany, the United States, and China (https://www.indexbox.io/blog/which-country-consumes-the-most-linseed-oil-in-the-world/).

Nowadays nutritional aspects of linseed are gaining more attention from nutraceutical and pharmaceutical industries. The phytochemical components present in seeds of flaxseed are α-linolenic acid (ALA) or omega-3 fatty acid, linoleic acid, oleic acid, digestible proteins, carbohydrates, lignans (phytoestrogen), soluble mucilage, and dietary fiber and an array of antioxidants. Approximately, 21% and 34% of proteins are found in flaxseed grain and ground flaxseed paste, respectively (Chung et al. 2005).

Flaxseeds are a good source of sulfur-containing (methionine and cysteine) and branched-chain (isoleucine, leucine, and valine) amino acids. It also comprises limiting amino acids like lysine, threonine, and tyrosine. Additionally, flaxseed is a good source of various kinds of phenolic acids such as ferulic acid (10.9 mg/g), chlorogenic acid (7.5 mg/g), and gallic acid (2.8 mg/g) along with some minor phenolic acids such as ρ-coumaric acid glucosides, hydroxycinnamic acid glucosides, and 4-hydroxybenzoic acid. Flaxseeds also constitute a rich source of minerals like calcium (236 mg/100 g), magnesium (431 mg/100 g), phosphorus (622 mg/100 g), and potassium (831 mg/100 g) and vitamins like γ-tocopherol (522 mg/100 g),

Table 1 Area, production and productivity of linseed in major producing countries

Country	Area (ha)	Production (tons)	Productivity (kg/ha)	Contribution to world production (%)
Kazakhstan	1,342,518	1,058,247	788	31.4
Russian Federation	973,411	787,923	809	23.4
Canada	371,000	578,000	1558	17.2
China	250,000	330,000	1320	9.8
India	200,000	121,000	605	3.6
United States of America	119,790	144,940	1210	4.3
Ethiopia	78,921	80,457	1019	2.4
Total world	3,540,139	3,367,331	951.2	–

Source: FAOSTAT (2020) (https://www.fao.org/faostat/en/#data/QCL)

α-tocopherol (7 mg/100 g), δ-tocopherol (10 mg/100 g), and niacin (3.2 mg/100 g) (Bernacchia et al. 2014). The seeds possess a pleasant nutty taste with a crisp and chewy texture (Carter 1993). Generally, consumption of one tablespoon (7 g) of ground flaxseed is recommended per day for a healthy person (Meacham et al. 2022). The bioavailability of nutrients is improved only after processing flax seeds. About 20 to 25% of daily fiber needs can be fulfilled by consuming half an ounce of dry whole flax seeds, and 30 g of flaxseeds contain 7 to 30% of the Recommended Dietary Allowances of calcium, magnesium, and phosphorus (Singh et al. 2011). The above nutritional components present in the seeds of linseed have a vital role in human diet and health maintenance.

Human diseases are caused by genetic, environmental, pathogenic, nutritional, food habits, and lifestyle reasons. Adding functional foods like nutrient-rich vegetables, fruits, seeds, grains, and nuts to the food will improve human health. Researchers proved that consumption of linseed have a potentially positive effect on health beyond basic nutrition and can prevent, postpone and besides being supportive in management of several kinds of diseases like cardiovascular complications, cancer, kidney disease, dry eye disease, polycystic ovary syndrome (PCOS), gout, diabetes mellitus, liver disorders, neurodegenerative disorders, stomach ulcers, constipation, etc. (Gaber and Badawy 2019; Rezaei et al. 2020; Saivarshine et al. 2020).

To harness the potential of linseed, there is a need to develop the varieties/hybrids with higher seed yield and resistance to pests and diseases along with environmental stress tolerance and higher nutritional values. Focus of conventional breeding is mostly on agronomic traits related with yield and fiber quality along with pest and disease resistance. There is an immense need to understand and improve the nutritional aspects of linseed cultivars. In this context, utilization of advancements in molecular and biotechnological strategies will pave the way for successful linseed breeding and its effective utilization in nutraceutical and pharmaceutical industries.

2 Nutritional Compounds

The linseed oil is the richest vegetarian source of α-linolenic acid (ALA; ω3-C18: $3^{9,12,15}$), which is also known as omega-3 fatty acid. Human body cannot synthesize omega-3 fatty acids by utilizing any other substrates and, hence, must be supplied through food. These kinds of fatty acids are known as essential fatty acids. ALA is the parent fatty acid of the ω3 family which includes stearidonic acid (SDA; ω3–18: $4^{6,9,12,15}$), eicosapentaenoic (EPA; ω3–20:$5^{5,8,11,14,17}$), and docosahexaenoic acid (DHA; ω3 22:$6^{4,7,10,13,16,19}$) (Fig. 1). These omega-3 fatty acids are crucial for optimal development of the human brain, circulating cells (red blood cells (RBCs), leukocytes, etc.), and skin (Ratnayake and Galli 2009). When flaxseeds are consumed, ALA may alter plasma oxylipin concentration which produces antihypertensive effects in patients suffering from peripheral arterial disease (Caligiuri et al. 2014).

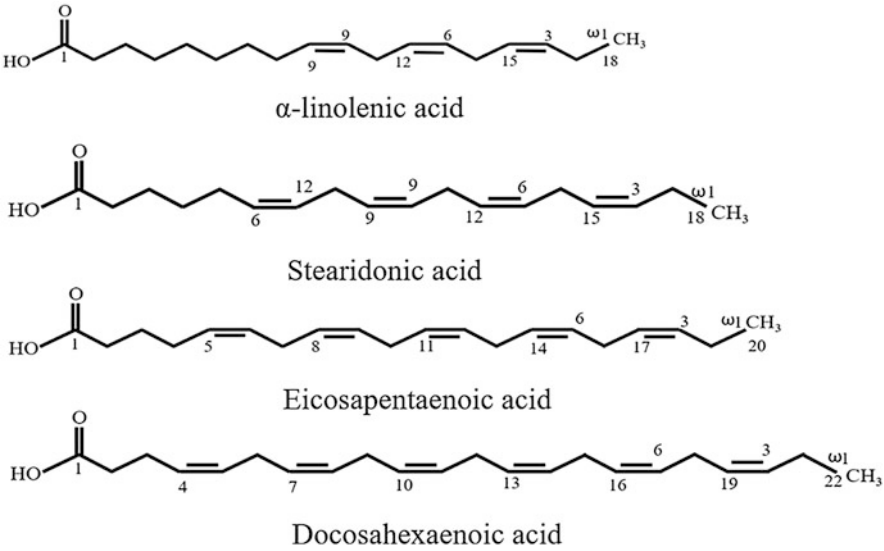


Fig. 1 Structures of omega-3 fatty acids

Flaxseed is one of the richest plant sources of lignans [secoisolariciresinol diglucoside (SDG- 294–700 mg/100 g), pinoresinol (3.32 mg/100 g), matairesinol (0.55 mg/100 g), and lariciresinol (3.04 mg/100 g)], providing 800 times more lignans than most other foods. When flaxseed is consumed, SDG lignans are metabolized in the intestinal region by gastrointestinal bacteria to enterolactone (the major active mammalian lignan that is found in body tissues) and also enterodiol (also a mammalian lignan). The enterolignans, enterolactone, and enterodiol are known to possess weak estrogenic activity. Because of this, SDG is classified as phytoestrogen. Weak hormonal action of flaxseed lignan competes with the same receptor for estrogen. This may decrease the hormonal signaling involved at the beginning of tumor development in the breast (Dyer 2014).

Intake of flaxseed lignans reduces the risk of polycystic ovary syndrome (PCOS) in susceptible women as lignan binds with free circulating testosterone and excluded the biliary system. Studies with flaxseed supplementation (30 g/day) for 4 months in a 31-year-old PCOS woman showed a reduction in total testosterone (from150 ng/dL to 45 ng/dL) and free testosterone (from 4.7 ng/dL to 0.5 ng/dL) along with reduced hirsutism (a condition of unwanted male-pattern hair growth in female) (Nowak et al. 2007).

Flaxseed has very low levels of carbohydrates (1 g/100 g). Due to high dietary fiber, higher levels of micronutrients, and lower levels of carbohydrate, flaxseed cake has low glycemic index. In a recent study, incorporation of flaxseed as 10 g in three servings to a daily diet in the form of muffins lowered glucose profile over 24 h

compared to the control (0 g of flaxseed) (Almehmadi et al. 2021). This study highlights the key role of flaxseed cake/flour for control and prevention of diabetes. Consumption of flaxseed protein hydrolysate prevents the production of potent blood vessel constrictor octapeptide angiotensin II by inhibiting the angiotensin-I-converting enzyme and thus reduces the blood pressure (Marambe et al. 2008). Among many reasons, nitric oxide an endogenously synthesized free radical within human body causes neuroinflammation and progress in Alzheimer's disease. Synthesis of nitric oxide requires transport of electrons from the carboxyl-terminal domain to the amino-terminal domain of nitric oxide synthase by calmodulin. Omoni and Aluko (2006) observed that flaxseed protein hydrolysate induced a change in secondary and tertiary structures of calmodulin which presumably decreases the production of nitric oxide and thus reduce the risk of neurodegenerative disorders.

Linseed mucilage is obtained from the seed coat. Aqueous extract of mucilage from seed by precipitation process (soaking in water for 24 hours) gives high viscosity solution that forms a white friable, fibrous mass when fully dry. Its functional properties are equivalent to gum Arabic, and its composition includes polysaccharides like xylose, glucose, galactose, galacturonic acid, arabinose, and rhamnose. Linseed mucilage is utilized as an emulsifier and thickener in food products, inflammation or irritation reliever in pharmaceutical industries, and a base material for eye ointment (Das and Naik 2017).

Utilization of nutritional compounds of linseed in nutraceutical and pharmaceutical industries is gaining momentum along with research findings in human health improvement by linseed products (Fig. 2). Flax seed lignan extract is available in the market as 40 mg capsules and recommended as a support for menopausal and postmenopausal women. Flax lignan SDG 156 mg of 60 capsules available in the market is used as a dietary supplement which is having strong antioxidant activity; promotes prostate, breast, circulatory, and bone health; and helps alleviate discomforts associated with menopause. Concentrated linseed SDG lignan 35 mg capsules are commercially available in the market and are mainly used for treating liver and hormonal problems in the dog.

Fortification of α-linolenic acid in meat, eggs, bakery, dairy, and infant formula is effective and an economical way to deliver omega-3 fatty acid among health-conscious consumers (Patel et al. 2022). Nanotechnology delivery system can be effectively used for fortification of food with bioactive lipids (fat-soluble vitamins, omega-3 fatty acids) to improve nutritional quality and the bioavailability of the products. Novel nano-emulsion for fortification of dairy beverages with α-linolenic acid and essential micronutrients, production of omega-3 yogurt using flaxseed oil, vegetarian omega-3 fatty acid enriched cake, linseed oil-based nano-encapsulation using chia seed mucilage, and ice cream using microencapsulated flaxseed oil are some of outcomes from nanotechnological approach (Gowda et al. 2018; Stefani et al. 2019; Jagtap et al. 2020; Almasi et al. 2021; Murugkar et al. 2021). It was found that microencapsulation not only improved antioxidant activity of food but also improved shelf life and sensory attributes (Eric-Parfait Kouamé et al. 2021).

- **Flaxseed (*Linum usitatissimum* L.)**
- **Major constituents:**
  - ❖ Flaxseed oil (Alpha-linolenic acid)
  - ❖ Dietary Fibre
  - ❖ Lignan - Secoisolariciresinol diglucoside (SDG)
  - ❖ Proteins
  - ❖ Carbohydrate
  - ❖ Amino acids
  - ❖ Vitamins
  - ❖ Minor micronutrients
  - ❖ Antioxidants

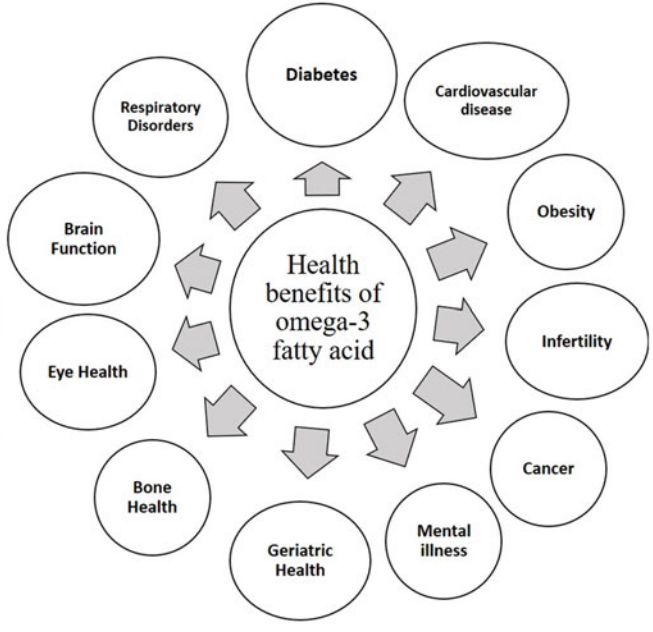


Fig. 2 Major constituents, health benefits, and diversified products of linseed

3 Genetic Resources of Health-Related (HR) Genes

Plant genetic resources (PGR) with a wider genetic base are the most important and essential basic materials to meet the current and future requirements of crop improvement programs. Wild relatives, breeding lines, elite cultivars, native landraces, local selections, mutants, polyploids, and hybrids of crop plants are essential components of PGR. The diversity in PGR is systematically collected and conserved by PGR institutes/centers. Around 48,000 linseed germplasm accessions are conserved in gene banks all over the world. Approximately 3500 accessions of world cultivated flax collection are being maintained at Plant Gene Resources of Canada (PGRC 2018). Approximately, 2900 linseed accessions are stored under a long-term storage facility of the National Bureau of Plant Genetic Resources (NBPGR) in India. The crop gene pool concept was first developed by Harlan and de Wet (1971). Further, they classified the germplasm resources into primary, secondary, and tertiary gene pools. Cultivated flax (*L. usitatissimum*) and its progenitor pale flax (*L. bienne*) are together considered as primary gene pool. Both are diploid species with $2n = 30$ chromosomes. They are cross compatible with each other in both directions and yield fertile progeny. Those closely related flax species which can be crossed with cultivated flax and can give partially fertile progeny are kept under secondary gene pool. Wild flax species like *L. nervosum*, *L. pallescens*, *L. africanum*, *L. corymbiferum*, *L. decumbens*, *L. hirsutum*, *L. floccosum*, and *L. tenue* produced successful cross progeny with cultivated flax, and these can be considered as the secondary gene pool. Those species which are distantly related to flax and cross incompatible with cultivated flax but could be exploited using advanced biotechnological tools like genetic engineering (Fu 2019) are classified as tertiary gene pool. Quaternary gene pool was introduced by Gepts and Papa (2003) which includes genetically engineered flax like CDC Triffid. This classification helps to assist the access of genetic resources of the crop for plant breeding and to assess the extent of gene flow among populations of a crop and related taxa. Genetic resources in different gene pools need to be expanded and evaluated to identify the germplasm with useful traits like pest and disease resistance, high oil content, high or low omega-3 acid and high or low seed mucilage content, higher thousand seed weight, higher seed yield, etc. Two similar types of linseed are commonly traded, brown and yellow (golden) seeds which are closely related in terms of oil content, but the brown seeds contain more alpha-linolenic fatty acids (59%) when compared to the yellow seeds (51%). Therefore, brown linseed can be considered as nutritionally richer than yellow seeds. Both types of linseed are used for human consumption. Mostly yellow linseed is considered for cooking purposes because it blends well with various dishes due to its golden color. Among the many traits, seed mucilage content and omega-3 fatty acid are important traits with huge nutraceutical values. Diederichsen et al. (2006) reported a wide variation of seed coat mucilage content in 1689 flax accessions from 61 countries based on mucilage indicator values (MIVs) ranging from 22.1 to 343.4 centistokes (cSt) mL g^{-1}. These accessions can be utilized for mapping quantitative trait loci (QTL) and genes that are associated with mucilage content by linkage analysis and association mapping.

4 Classical Genetics and Traditional Breeding

The genus *Linum* consists of approximately 230 species and separated into 5 sections, *Cathartolinum*, *Dasylinum*, *Linum*, *Linastrum*, and *Syllinum*, based on floral morphology, chromosome number, and interspecific compatibility. Cultivated flax, *L. usitatissimum*, and its suggested progenitors *L. angustifolium* (also known as pale flax) and *L. bienne* are placed under section *Linum*.

Improvement in flax has lagged behind other oilseeds because it occupies a smaller niche as an oilseed and, consequently, received limited resources for research and developmental efforts. Exploitable genetic diversity is also very low within the crop and cannot be readily supplemented by intraspecific hybridization. Hence, breeding strategies in flax involve creating additional genetic variability, selecting the best recombinants, and fixing the genes by making them homozygous. Inheritance of crop characteristics can be controlled by one major gene not influenced by the environment (qualitative traits) to many genes, much influenced by the environment (quantitative traits). Since flax is an autogamous crop, many important traits in flax including days to flowering, days to maturity, plant height, branching and seed yield are quantitative in nature and demonstrate additive gene effects (Mohammadi et al. 2010).

The traditional breeding methods of flax are mainly concentrated on increasing and stabilizing seed yield, enhancing oil content and quality, increasing seed size, and discovering resistance sources to major diseases, early maturity, lodging resistance, plant height, etc. (Dhirhi et al. 2018; Terfa and Gurmu 2020). The selection of genetically diverse parental material in flax breeding depends on the interested trait(s), the crossing purpose, characters which are relatively important other than yield, the pedigree of the pure lines, and the resources and time available. The parental lines can be from many different sources including existing cultivars, adapted elite breeding lines, or new introductions from global flax collection. Flax breeding methods based on inbreeding selection and line evaluation include pure line method, pedigree method, bulk population method, single-seed descent, backcross breeding, recurrent selection, and early-generation testing. Other methods like male sterility, mutation breeding, molecular marker-assisted selection, haploid culture, anther culture, and genetic engineering have much greater potential to enhance and accelerate the flax breeding programs. Commercially cultivated flax cultivars are pure lines which are developed through hybridization of inbred lines along with pedigree selection of recombinant lines from segregating populations (Hall et al. 2016). Consequently, cultivated flax programs continue to rely on a narrow genetic base.

Wang et al. (2017) evaluated six different cultivars of oil and fiber types and evaluated phytochemical compounds, namely, caffeic acid, ρ-coumaric acid, and ferulic acid, and secoisolariciresinol diglucoside (SDG) along with total phenolics, flavonoid contents, and total antioxidant activities. Phytochemical profile of oil- and fiber-type cultivars was non-significantly altered, while cellular antioxidant activity was more in fiber-type cultivars. This study indicated fiber-type cultivars can also be suitable option as that of oil-type cultivars for functional products and dietary supplement production.

Recently, Zhang et al. (2022) studied the effects of environment and genotype along with their interactions on lignan (SDG) content of flaxseed. In this study, six genotypes, namely, Zhangya 2, Longya 8, Linxiabai, Gaolanbai, Shandanbai, and DYMS, were evaluated at eight different locations in China. The lignan (SDG) content in flaxseed ranged between 4.71 and 7.27 mg/g, and this study revealed that among the total recorded variation, 51.38% was due to genotypes and 44.40% due to environments. Further altitude had positively affected lignan content. The study concluded SDG content is mainly genetically controlled and essentially affected by environmental conditions.

Mutation-based breeding, using radiation or chemical mutagenesis, is useful to upgrade elite lines by altering one or two production-limiting or quality traits. Modification of fatty acid profiles, reduction of bast fiber content (McKenzie 2011), and development of new flax type with edible flaxseed oil by eliminating α-linolenic acid (ALA) (18:3cis$\Delta^{9,12,15}$) is also possible through mutation breeding. Mutation breeding has the possibility to remove undesirable compounds from flaxseed, such as linatine and cyanogenic glycosides (Green et al. 2008). Most interesting low and medium linolenic acid linseed lines were obtained through use of ethyl methanesulfonate (EMS) (Ntiamoah and Rowland 1997). Time-consuming procedures, the lack of creation of variability, and narrow genetic base of developed varieties are considered as limitations in conventional breeding. However, molecular methods are being developed to support and speed up the conventional breeding programs.

5 Genetic Diversity Analysis

Important utilizations of genetic diversity analysis (GDA) are an examination of evolutionary relationships of wild species, development of the genetic-linkage and physical mapping, QTL mapping, association mapping, estimation of genetic distance between different accessions, identification of novel trait-specific alleles and discovering microsatellite markers, fingerprinting and cultivar identification, etc. GDA provides an option for the selection of new breeding lines in sustainable plant breeding program. Morphological parameters and isozyme markers were first used for the determination of genetic diversity in linseed. Two distinct morphological types, namely, flax and linseed, were recognized in cultivated linseed species. The flax-type cultivars are mainly grown for the extraction of fiber and are tall-growing with straight stems and few numbers of secondary branches. The linseed types were shorter with lots of branches and grown for seeds, and those predominantly grown in India are meant for extraction of the oil. Diversity analysis based on morphology in flax includes characteristics like plant height (cm), petal color, days to flowering, days to maturity, capsules per plant, seeds per capsule, seed color, seed yield per plant, thousand seed weight (g), oil content (%), oil yield per plot, etc. Adequate variability among the genotypes was identified by GDA, and available genotypes were classified into different groups based on cluster analysis. Genotypes under different clusters with maximum distance can be exploited in the linseed improvement programs (Patial et al. 2019; Kumar and Kumar 2021). However, morphological characteristics are highly

affected by the environment and require labor-intensive field evaluation over an extended period. In addition, biochemical markers are affected by plant developmental stages. Hence, these limitations with morphological and biochemical markers led to the utilization of DNA-based markers for GDA.

Oh et al. (2000) first reported DNA-based markers utilization in linseed for assessing the genetic diversity study. Molecular markers are a specific fragment of DNA sequences found at particular locations on the genome which shows polymorphism between dissimilar individuals and are transmitted from one generation to the next by the standard laws of inheritance. Molecular markers identify insertions, deletions, point mutations, or inversions that occur in allelic DNA sequences, which can distinguish every individual in the same species (Prabha et al. 2017). Various types of molecular marker systems including random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), inter-simple sequence repeat (ISSR), inter-retrotransposon amplified polymorphism (IRAP), expressed sequence tag (EST), and simple sequence repeat (SSR) have been utilized for GDA of flax germplasm. DNA-based molecular characterization has numerous advantages like abundance, non-tissue-specific traits, early and rapid assessment, and environment independence and can be effectively utilized in GDA (Nag et al. 2015).

The genetic diversity available for mucilage content in flax germplasm makes it feasible to create breeding approaches for high seed mucilage (SM) flax for human consumption and low SM flax for animals and fish. Soto-Cerda et al. (2012) used 150 microsatellite loci to evaluate the linkage disequilibrium (LD), population structure, and genetic diversity of 60 linseed cultivars/accessions covering the breadth of seed mucilage variation in linseed germplasm. Similarity-based methods and STRUCTURE analysis showed the presence of populations expressing particular their geographic origins (North America, South America, and South Asia) and the influence of germplasm exchange among North American flax breeding programs. The spatial distribution of the accessions in the principal coordinate analysis (PCoA) diagram revealed the occurrence of higher genetic distance among the South Asian population and more similarity within the genetic composition of the South American and North American populations. The genetic diversity for each population showed the presence of higher polymorphism information content (PIC) values, allele numbers, and alleles per locus in the North American and South American populations than in the South Asian population. Thirty accessions, covering the allele richness of 99.5% and a comparable range of mucilage variation that exists in the sixty accessions were identified as the ideal core set, which helps to optimize the detection of significant associations between SSR alleles and mucilage-related genes/QTLs in flax. The results conclude that the collection of flax accessions could be effectively utilized in association mapping (AM) studies targeting the discovery of alleles/genes involved in seed mucilage. However, there is a need for greater diversity to improve the AM resolution in flax.

Maximization of genetic gain in linseed needs increasing of genetic diversity of linseed parental stocks by including diverse genotypes. For this purpose, Hoque et al. (2020) analyzed the population substructure, linkage disequilibrium, and genetic diversity of 350 linseed genotypes distributed globally by utilizing 6200 single

nucleotide polymorphism (SNP) markers. The average marker density of SNP markers across the 15 chromosomes was 1 per 51.17 kb. The presence of more transitions and less transversions SNPs in linseed indicates more tolerable nature of transition SNPs in natural selection (Luo et al. 2017). Structure analysis using the PCoA, delta K approach, and neighbor joining tree analysis categorized the whole linseed genotype collections into seven sub-populations (P1 to P7). The genotypes that originated from Asia (India, Pakistan), Europe (Hungary), and Turkey were grouped under P6, P1, and P7, respectively, whereas North Dakota State University (NDSU) along with other American genotypes were categorized under P5 sub-population. P2, P3, and P4 consisted of a mixture of linseed genotypes from different origins. Oil-type genotypes were observed in all the sub-populations except P2, which included mostly the fiber-type linseed genotypes. Among oil types, P1 and P7 sub-populations belonged to winter type. Spring-type seed flax belonged to P5, Indian seed flax with short large seed belonged to P6, the Argentine or Mediterranean seed flax belonged to P3 and forage type, and seed flax of Ethiopia belonged to P4 sub-population. The identified SNP markers revealed only moderate diversity in different sub-populations. Negative Tajima's D value of sub-population P6 indicated occurrence of more rare alleles in P6 group, and positive Tajima's D values of remaining six sub-populations indicated less rare alleles in groups other than P6. This concluded P6 as a group with recent expansion and remaining as groups of recent population contraction. Hybridization between Asian individuals (P6) and NDSU stock (P5) is possibly the best choice to increase genetic gain. Dual-purpose varieties could be developed by crossing P2 and P5 because of their low genetic differentiation and possibilities for accumulation of desired alleles for oil and fiber content.

Though cultivation of linseed has been practiced over a period of 1000 years or more as an oil and fiber crop, it was unclear about the changes in agronomic traits at the genetic level during linseed cultivation. Zhang et al. (2020) performed whole-genome shotgun sequencing on multiple accessions of oil-use, fiber-use, landraces, and pale flax to identify the genomic variations during flax cultivation. The findings of this study revealed that the genes pertinent to plant architecture, flowering, dehiscence, and oil production were favorably selected during linseed domestication. Furthermore, regardless of the origin, the advancement in modern oil-use flax is more compared to fiber-use flax, although the selection for both oil and fiber flax occurred simultaneously during the early period of flax domestication. They also found that the spread of *MYB46/MYB83* sister genes in flax may cause the biosynthesis of unique secondary cell wall in flax stem. The directional selections of *MYB46/MYB83* may have provided a shape to morphological profile of the present oil and fiber flax.

It is well identified that genes of fatty acid desaturase (*FAD*) and stearoyl-ACP desaturase (*SAD*) families play an important role in fatty acid synthesis, and few alleles of these genes are inked with flax oil synthesis. Genetic polymorphism of *FAD* and *SAD* genes was assessed in the representative accessions of 84 linseed cultivars and lines showing different levels of palmitic (5–7.6%), stearic (2.7–6.4%), oleic (12.9–24%), linoleic (11.9–72.4%), and linolenic (2.7–65.3%) acids. The highest level of genetic diversity was observed for *FAD3A* and *FAD3B* genes. The polymorphisms identified in this study could be effectively utilized for the marker-assisted selection in flax breeding (Dmitriev et al. 2020).

6 Molecular Mapping of HR Genes and QTLs

Association mapping (AM) approach is effectively applied for identification of QTLs by examining the marker-trait associations across the set of diverse germplasm/population. Identification of alleles and new genes as well as dissection of complex characters is possible by AM. Genome-wide association studies (GWAS) and candidate genes are important components of AM. The selection of germplasm, phenotypic and genotypic data quality, and use of suitable statistical analysis for detection and verification of marker-phenotype associations are critical to association analysis. AM provides a high-resolution than the linkage map to determine the importance of various loci or segment of the genome on the expression of single or multiple traits, because AM utilizes an extensive range of germplasm including natural populations and collections of varieties and breeding lines to map traits by linkage disequilibrium (LD). Population structure, population and sample sizes, and marker density are few critical factors which should be considered in application of AM (Ibrahim et al. 2020). Cloutier et al. (2009) reported the development of 275 polymorphic EST-SSRs in flax which could be used to develop physical maps, QTL mapping, and potential to reduce the time for cultivar development. Following the development of these markers, a consensus genetic and physical map using DNA markers in combination with the flax genome sequencing was reported (Cloutier et al. 2012). Three dissimilar mapping populations (CDC Bethune/MacBeth, E1474/Viking, and SP2047/UGG5–5) were used to generate this map. The consensus map has 670 DNA markers anchored to 204 of the 416 fingerprinted contigs of the physical map, and the overall map density is 2.0 cM for markers arranged on 15 linkage groups.

Linseed oil is used for its food and non-food applications. Modification in oil content along with fatty acid profiles of linseed to fulfill the needs of the market promptly requires a clear understanding of their QTL architectures, which have received poor attention to date. Soto-Cerda et al. (2014) assayed a flax core collection of 390 accessions with 460 microsatellite markers for the quantitative nature of seed quality traits including oil content (OIL), palmitic acid, stearic acid, oleic acid, linoleic acid (LIO), linolenic acid (LIN), and iodine value. After growing the core collection in a modified augmented design at two locations over 3 years, phenotypic data for all the seven traits were obtained from six environments. Substantial phenotypic diversity was noticed for each trait along with moderate to higher heritability. Most of the stable candidate QTLs were identified by multivariate analyses. Co-localizations of candidate QTLs for LIO and LIN were observed with previously identified QTL of bi-parental populations. Some QTLs were mapped with the genes which are considered to be responsible for fatty acid biosynthesis. Canadian cultivars showed absence of 58 percent of QTL alleles when compared to core collections which indicated potential usefulness of core collection in improvement of seed quality traits. The candidate QTLs identified in the particular study laid the foundation for marker-assisted breeding studies in linseed.

Kumar et al. (2015) constructed a genetic map based on selected 329 SNPs and 362 SSR markers using a recombinant inbred line population of 243 individuals from a cross between the Canadian varieties CDC Bethune and Macbeth. The

genetic map comprised of 15 linkage groups including 691 markers where one marker in every 1.9 cM was the average marker density. Total number of QTLs identified in this study with respect to 14 traits was 20. Among them, two QTLs for linoleic acid, one QTL each for linolenic acid, palmitic acid, and oil content, and three QTLs each for stearic acid and oleic acid were identified. Proteins synthesized by candidate gene regions linked with QTLs were involved in yield component traits along with fatty acid biosynthesis and their metabolism. This study provided a base for map-based QTL cloning and marker-assisted selection (MAS) in linseed.

Soto-Cerda et al. (2012) assessed 62 flax cultivar/accessions to understand the LD, population structure, and genetic diversity of linseed germplasm showing variation in mucilage content. It is crucial for the recognition of QTLs and genes related to mucilage by association mapping.

GWAS of flaxseed was conducted by Soto-Cerda et al. (2018) for mucilage content and hull content. A total of 200 diverse linseed accessions were genotyped by using 1.7 million SNP markers. The data revealed a wide range of phenotypic variation along with high (~70%) and an average (~49%) narrow-sense heritability for mucilage content and hull content, respectively. Seven QTLs for MC and four QTLs for HC were also identified. Promising candidate genes such as *TRANSPARENT TESTA 8*, *GALACTUROSYL TRANSFERASE-LIKE 5*, *SUBTILISIN-LIKE SERINE PROTEASE*, *MUCILAGE-MODIFIED 4*, *AGAMOUS-LIKE MADS-BOX PROTEIN AGL62*, *GLYCOSYL HYDROLASE FAMILY 17*, and *UDP-GLUCOSE FLAVONOL 3-O-GLUCOSYLTRANSFERASE* were identified which are *Arabidopsis thaliana* gene orthologs to *L. usitatissimum*. These genes play a significant role in seed coat development, synthesis, and release of mucilage and biosynthesis of anthocyanins in *A. thaliana*.

Several software/statistical packages have been utilized in AM studies. These include Trait Analysis by Association, Evolution and Linkage (TASSEL), STRUCTURE, R package, EINGENSTRAT, Statistical Analysis System (SAS), Spatial Pattern Analysis of Genetic Diversity (SPAGeDi), Multiple Trait Derivative-Free Restricted Maximum Likelihood (MTDFREML), Residual Maximum Likelihood (ASREML), STRAT, Bimbam, and GEN STAT 11 (Álvarez et al. 2014). Heterogeneity in dependent instruments (HEIDI) and Summary-data-based Mendelian randomization (SMR) tools are utilized to test pleiotropic interaction between gene expression level and complex traits using expression quantitative trait loci (eQTL) and GWAS data (Zhu et al. 2016).

7 Marker-Assisted Breeding for HR Traits

Marker-assisted breeding (MAB) combines classical plant breeding with the tools and discoveries of molecular biology and genetics, most specifically the utilization of molecular markers to develop new varieties/hybrids in cultivated crops. Utilizing molecular markers in linseed improvement started in the late 1990s. Initially RAPD and RFLP markers were developed along with isozymes and morphological markers for construction of genetic maps. The discovery of PCR-based marker system

provided new opportunities to identify several DNA polymorphisms within linseed genomes. Among many PCR-based markers, SSRs/microsatellite markers based on genomic structural variation became popular because of their highly polymorphic and robust nature and relatively simple and inexpensive analysis. In earlier study, Cloutier et al. (2009) developed 275 EST-SSR loci and studied the genetic relationships of 23 linseed accessions. Two microsatellite-enriched linseed genomic libraries were constructed by Deng et al. (2011) for trinucleotide ATC and TTC motifs. They also characterized and evaluated 38 polymorphic microsatellite markers in 8 linseed cultivars from different regions of the world. These research findings are useful in identification and classification of germplasm, construction of genetic linkage map, QTL mapping along with gene identification, and marker-assisted breeding of linseed. Wu et al. (2017) screened 1574 microsatellites from linseed obtained using reduced representation genome sequencing (RRGS) to systematically identify SSR markers that would be immediately suitable for use in flax breeding. The identified SSR markers were immediately utilized to distinguish oil and fiber types of flax among the 48 varieties screened. 'NEW1' and 'Venus' exhibited distinct genetic backgrounds from the other fiber cultivars. Similarly, cultivar 'A0529' showed a different genetic composition than the other oilseed cultivar. Therefore, it was concluded that these three cultivars, 'NEW1', 'Venus', and 'A0529', have the huge potential to use in flax improvement programs.

SNPs which are highly abundant and known to be present in high frequency in the genome gained attention during the past decade. Kumar et al. (2012) discovered 55,465 SNPs using reduced representation libraries of 8 linseed genotypes and validated 4706 SNPs using genotyping-by-sequencing of 96 F_6 individuals from a population of recombinant inbred lines derived from a cross of CDC Bethune and Macbeth. The authors recommended SNP marker developed in this work for construction of high-density maps of linseed, association mapping, phylogenetic analyses, QTL discovery, MAS, and whole genome shotgun sequencing. Advancements in molecular marker development resulted in their effective utilization in construction of high-density linkage map (Yi et al. 2017), genomic variations and association study (Xie et al. 2018), and identification of powdery mildew resistance source (Speck et al. 2022) in linseed through marker-assisted breeding.

8 Genomics-Aided Breeding for HR Traits

Breeders can select suitable parents with novel combinations for different crossing programs by genomics-assisted breeding (GAB) which ultimately helps to develop the elite breeding lines. Adequate genomic resources coupled with broad genetic base in cultivated gene pool are a prerequisite for GAB. Several research groups are engaged in developing genomic resources during the last few years in linseed.

Advancement in molecular genetic methods led to development of numerous molecular markers allowing breeders to use the markers to aid breeding. Though MAS was proposed to improve the linseed traits that are governed by QTLs or genes with relatively large effects, most of the economically important traits of crops are

controlled by polygenes, each with a smaller effect (Riedelsheimer et al. 2012). This is a limitation of MAS in linseed breeding. Therefore, an advanced method, i.e., genomic selection (GS) which enables to evaluate breeding values (BVs) of the segregating individuals of crosses, was introduced. GS is considered as an advancement of MAS. In GS, a statistical prediction model is built to predict breeding values (BVs) called genomic estimated breeding values (GEBVs) of unphenotyped individuals by using genome-wide markers in a genotyped and phenotyped training population (Meuwissen et al. 2001). This is highly useful in breeding to predict GEBVs of individuals and thus identify superior genotypes among selection candidates according to their genomic information.

In another advancement, the integration of GS and computer simulation resulted in an improved genomic tool known as genomic cross prediction. Genomic cross prediction uses the available information of genome-wide molecular markers and consensus genetic maps and integrates computer simulation and GS to simulate virtual crosses, generate segregation populations, and estimate the expected genetic performance of the virtual crosses to assist breeders in selecting parents and crosses in plant breeding effectively.

You et al. (2022) extracted a total of 290 linseed accessions from the flax core collection as a training population. These accessions are possible parents in flax breeding, including 13 landraces, 193 cultivars, 59 breeding lines, and 25 unknown lines collected from 34 countries. A set of 258,708 SNPs in 290 flax accessions were selected (He et al. 2019) for further analyses. Five important linseed traits including days to maturity (DTM), seed yield (YLD), linolenic acid (LIN), oil (OIL), and powdery mildew resistance (PM) were chosen as breeding targets in flax breeding selection. Genomic selection models were constructed, virtual crosses and simulation of progeny populations, evaluation of virtual crosses were carried out. A total of 317, 450, 313, 496, and 235 nonredundant QTNs were identified using six multi-locus models for DTM, YLD, LIN, OIL, and PM, respectively. The predictive abilities for all five traits were greater than 0.90, being 0.95, 0.90, 0.95, 0.92, and 0.89 for YLD, DTM, OIL, LIN, and PM, respectively. GCAs of 290 parents were calculated for all 41,905 single crosses. A consistently high linear relationship, i.e., close to 1, was observed between GCAs and GEBVs of the parents, suggesting that GEBVs of parents estimated using GS models with QTN markers can effectively predict the GCAs of the parents. In linseed breeding programs, early maturing, disease-resistant individuals with high YLD, OIL, and LIN were selected. The top 10% accessions (29 out of 290 accessions) with the highest GCAs for each trait, which have the best potential to improve traits in linseed cross breeding, were identified.

9 Cloning of HR Genes

Industrial oil flax and edible oil or solin flax differ mainly in seed linolenic acid levels. Seed oil of flax contains high level of linolenic acid, generally around 45% to 65%. Rowland (1991) observed an M_2 seed with mutations that contain less than 2% linolenic acid by using chemical (EMS) mutagenesis. Two genes, *LuFAD3A* and *LuFAD3B*, are responsible for high linolenic acid in wild flax. Microsomal

desaturases encoded by these two genes are capable of desaturating linoleic acid. Vrinten et al. (2005) reported that cloning of two members of the FATTY ACID DESATURASE (FAD3) family of microsomal Δ15-FADs, designated *LuFAD3A* and *LuFAD3B*, reveals the molecular mechanism behind this trait. The expression of these two genes in *Saccharomyces cerevisiae* revealed point mutations in both the genes which result synthesis of premature stop codons and production of truncated inactive proteins in the low-linolenic-acid solin line.

A project on the Total Utilization Flax GENomics (TUFGEN; http://www.tufgen.ca) was started in Canada during 2009 to create genomics resources for linseed and to develop an inclusive knowledge of its unique genome with specific goals in applied genomics targeting flax improvement. In this study, genome-wide physical map for flax was constructed first time by using 43,776 bacterial artificial chromosome (BAC) clones from the library of the flax cultivar CDC Bethune. It provides a framework for accessing target loci with economic importance for marker development and positional cloning (Ragupathy et al. 2011).

Enzymes of *FAD2* and *FAD3* are responsible for the Δ12 [converted oleic acid (OLE) to linoleic acid (LIO)] and Δ15 desaturation [converted LIO to α-linolenic acid (LIN)] in planta. Three alleles of *FAD2A*, four of *FAD2B*, six of *FAD3A*, and seven of *FAD3B* were cloned in to a pYES vector (yeast expression vector), and all the 20 constructs were transformed in yeast along with an empty construct. The transformants were activated in the presence of substrates like OLE acid and LIO acid for FAD2 and FAD3, respectively. Fivefold increase of FAD3A-C activity and fourfold increase of FAD3A-F activity were identified. Utilization of such isoforms in breeding line development could enhance the relative amount of LIN acid in flax (Radovanovic et al. 2014).

Lysophosphatidic acid acyltransferases are an emerging family of enzymes that catalyze the synthesis of phosphatidic acid (PA) using lysophosphatidic acid (LPA) and acyl-CoA (Korbes et al. 2016). The product of this enzymatic reaction, PA, is precursor for the biosynthesis of all glycerophospholipids and triacylglycerol (TAG). Fahs et al. (2019) investigated the contribution of 2-lysophosphatidic acid acyltransferase (LPAAT) enzymes from flax in the accumulation of α-linolenic acid into the oil fraction of flax seed. cDNAs encoding three class A microsomal LPAAT2 isoforms were isolated from developing seeds of flax, and their specific activity was assayed. LPAAT2A showed high selectivity for linolenic and linoleic acids when compared to saturated fatty acids. Increased accumulation of linolenic and linoleic acids was observed in Arabidopsis seeds because of heterologous expression of LPAAT2A in two transgenic Arabidopsis lines. These findings suggest the incorporation of LPAAT2 isoform in linseed breeding program for developing cultivars with higher α-linolenic acid.

10 Genetic Engineering for HR Traits

Genetic engineering is considered as an excellent technology to enrich the gene pool of linseed/flax with genes of interest which are not naturally present in the linseed genome. The first genetically engineered flax cells were created in 1983 by

recombinant DNA technologies. Different methods to deliver genes into flax genome have been developed and utilized over the years. The most used procedure in the flax has been *Agrobacterium*-mediated hypocotyl transformation. Gene transfer using protoplast transformation, particle gun (biolistic) method (University of Saskatchewan, Saskatoon, Canada), *Agrobacterium* transformation of flax via floral dip (Dutta et al. 2021), and apical meristem-targeted in planta transformation protocol (Kesiraju et al. 2021) are also other successful methods to obtain transgenic flax. The fatty acid content and composition, the lignan (especially secoisolariciresinol diglucoside – SDG) content, flax fiber quality (improved elastic properties, reduction of pectin content, higher retting efficiency, improved mechanical properties, biodegradable polymers), tolerance to herbicides (sulfonylurea, glyphosate, and glufosinate-ammonium), and diseases resistance (*Fusarium*) have already been altered or improved by genetic engineering in flax (Rui-López et al. 2009; Ludvíková and Griga 2015). In 1996, the transgenic linseed cultivar CDC Triffid (sulfonylurea resistance) was officially registered in Canada, and it was deregistered in 2001 due to problems in its export to European countries. *Solanum sogarandinum* glycosyltransferase (SsGT1) gene was introduced into the flax genome to overproduce SsGT1 which is responsible for higher resistance to *Fusarium* infection in transgenic flax than wild-type plants. Overproduction of glycosyltransferase in transgenic flax also resulted in the accumulation of ferulic acid, ρ-coumaric acid, and caffeic acid and their glucoside derivatives, kaempferol, quercetin, and secoisolariciresinol diglucoside (SDG) in flax seeds (Czemplik et al. 2012). This transgenic flax seedcake extract showed a superior effect on fibroblast migration after a 24-hour treatment and inhibitory properties toward two bacterial strains: *Staphylococcus aureus* and *Escherichia coli*. This flax seedcake extract is good for skin health of human and can be considered an alternative treatment for infected wounds (Czemplik et al. 2012). Lorenc-Kukuła et al. (2005) aimed at generating flax plants with increased antioxidant capacity to increase the yield and resistance to *Fusarium culmorum* and *Fusarium oxysporum* via the expression of genes encoding chalcone isomerase, chalcone synthase, and dihydroflavonol reductase. Along with increase in antioxidant properties, pathogen resistance, and seed yield, increased content of SDG was also observed in transgenic lines. Introduction of *crtB*, a bacterial phytoene synthase gene, into flax by *Agrobacterium*-mediated transformation increased phytoene, α-carotene, and β-carotene content in seeds. The flow of phytoene production from geranylgeranyl diphosphate was first stimulated by the expressed *crtB* gene product in flax seeds. Then phytoene was further decomposed to α-carotene, β-carotene, and lutein, as catalyzed by the enzyme which biosynthesize endogenous carotenoid in seeds. These transgenic seeds expressed 65.4 to156.3 μg/g of carotenoid on fresh weight basis. A 7.8- to 18.6-fold enhancement of carotenoid content was observed in transformed seeds than in untransformed controls. These carotenoid-enriched transgenic seeds were considered as nutritional sources for human health (Fujisawa et al. 2008). Transformation of genes encoding acyl-desaturases and acyl-elongases in linseed from *Phaeodactylum tricornutum* (algae) and *Physcomitrella patens* (mosses) resulted in higher accumulation of 20C PUFAs including arachidonic acid (ARA) and eicosapentaenoic acid (EPA) in seeds of

transgenic flax (Abbadi et al. 2004). Introduction of the omega-3-specific D6-desaturase gene from *Primula vialii* in linseed by *Agrobacterium*-mediated transformation accumulates stearidonic acid (SDA) in flax seed. SDA is a metabolic intermediate on the biosynthetic pathway of eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), generated by the D6-desaturation of omega-3 fatty acid. Evidence suggests that SDA provides similar health benefits like EPA and DHA (Rui-López et al. 2009). The suppression of the chalcone synthase gene in transformed linseed plants resulted in the increased antioxidant status of seeds. Consequently, 20–45% of higher polyunsaturated fatty acid (mainly α-linolenic acid) accumulation was observed in transgenic seeds than in the control (Zuk et al. 2012). Pinoresinol lariciresinol reductase (PLR) is a key enzyme for SDG formation in the seed coat of flax seeds (Hano et al. 2006). RNAi phenomenon was successfully used to downregulate *LuPLR1* gene expression which leads to failed accumulation of SDG in plants. Synthesis of 8–5'linked neolignans dehydrodiconiferyl alcohol glucoside (DCG) and dihydro-dehydrodiconiferyl alcohol glucoside (DDCG) deviate the monolignol flux in linseed. These PLR devoid transgenic lines were utilized for the comparison of their behavior in the domain of insect resistance (Renouard et al. 2014). High oleic oil is a trait preferred by industries because of its stability and multiple applications. Two isoforms encoded by two *FAD2* genes in flaxseed, desaturate monounsaturated oleic acid to polyunsaturated linoleic acid. High content of oleic acid in flax, up to 80%, was realized in RNAi-mediated gene silenced lines by simultaneous silencing of both the FAD2 genes (Chen et al. 2015).

In general, plant oils do not contain very-long-chain polyunsaturated fatty acids (VLCPUFA). VLCPUFA play a major role in human physiology, either as elements of membrane phospholipids in specialized tissues or as precursors to produce the different groups of eicosanoid effectors. The nutritionally most essential VLCPUFA like arachidonic acid (ARA; ω6–20:$4^{5,8,11,14}$), EPA, and DHA are mostly available in fish oil. Synthesis of VLCPUFA in plant seed is possible by genetic engineering strategies. Seed-specific expression of cDNAs encoding fatty acyl-desaturases and elongases in transgenic linseed resulted in the very high accumulation of Δ6-desaturated C18 fatty acids and up to 5% of C20 polyunsaturated fatty acids (Abbadi et al. 2004). These results encourage biotechnological intervention on nutritional enchantment in linseed crop.

11 Social, Political, and Regulatory Issues

Nowadays food fortification with omega-3 fatty acid is gaining commercial importance. Any fortified food generally offers prophylactic effect and not therapeutic effect, and hence, a lot of time is required to understand the health benefits of any fortified products including omega-3 fortified products. Although health benefits of omega-3 fatty acids are well established, there is no fixed recommended daily allowance for omega-3 fatty acid. Further technological and regulatory issues

relating to food fortification are yet not fully clear with special reference to the appropriate levels of such ingredients in various foods (Panse and Phalke 2016).

Cadmium, cyanogenic compounds, and protease inhibitors present in seeds of linseed are toxic to human health. These are eliminated or reduced when linseed seeds are subjected to thermal treatments like frying in a pan, cooking in microwaves, autoclaving, and boiling (Austria et al. 2008). Some studies showed that flaxseed antinutrients have lesser effects when compared to soybean and canola. The consumption of flaxseed is not advisable in pregnancy, lactation, bleeding disorder, low blood pressure, bipolar disorder, food allergy, gastric obstructions, etc. (Ganorkar and Jain 2013). Daily consumption of three muffins (a total of 32.7 g of flaxseed) for 4 weeks didn't show any toxic or deleterious effect on the hemopoietic system or renal and hepatic function. However, serum triglyceride level was elevated (Stuglin and Prasad 2005). Though transgenic approaches showed lot of potential, strict legislative/political measures of European Union prevent the access of transgenic flax to the EU countries (Ludvíková and Griga 2015).

12 Future Prospects

Linseed improvement requires wider genetic variability, and it can be created with the aid of advanced molecular techniques. Application of marker-assisted breeding and genomic-aided breeding is the need of the hour to harness the nutraceutical and pharmaceutical potential of linseed crop. Flax has a long history of traditional use both as an oil and fiber source. Nowadays, there has been an emerging interest toward the nutraceutical and pharmaceutical properties of flax seed and its health benefits to human population. Though flax oil is the richest source of ALA, it is highly susceptible to oxidation and therefore has a very poor shelf life. Flax oil is generally cold-pressed and enriched by adding vitamin A and E or stored in dark glass jars to avoid faster rancidity. Since none of these protection methods are fully satisfactory, there is a necessity of appropriate breeding strategies, biotechnological interventions, and genetic engineering approaches for overproduction of various natural antioxidants within flax grains. Though conventional breeding and advancements in breeding approaches gave attention on agronomic traits and nutritional factors of linseed, there is a huge need to focus on reduction of toxic substance present in flax seeds. Genetic engineering approaches through transgenic route to create additional variability for targeted traits are available. In view of the regulatory concern, genome editing techniques like transcription activator-like effector nucleases (TALENs), zinc finger nucleases (ZFNs), and clustered regularly interspaced short palindromic repeats/CRISPR-associated (CRISPR/Cas) application have to be explored for linseed improvement. With about 48,000 cultivar accessions including 230 species belonging to primary, secondary, and tertiary gene pools, characterization of the entire set for exploitable variability for nutraceutical traits is yet to be accomplished. Unlike several other crops where the development is through research consortia with huge investments, linseed crop received very less attention and resources. Efforts need to be intensified to accelerate the breeding programs as

linseed with varied uses as food, fiber, and oil has great potential and a huge impact on human health, industry, and economy.

13 Conclusion

Linseed is getting huge demand considering its nutritional and medicinal properties. In order to fully exploit the nutritional potential of the linseed crop, technological advancements both in pre- and post-harvest is need of the hour. In this context, for crop improvement various crop-oriented technologies such as marker-assisted breeding and genomic-aided breeding, genome editing techniques followed by biotechnological interventions, and genetic engineering approaches for improved nutritional content and reduction of anti-nutrients in the crop for improving the edible use of flaxseed should be the right approach. Further, to utilize the medicinal and therapeutic potential of the crop, value addition using various post-harvest technologies for developing linseed-based fortified food products for improved human health and exploring its potential for food and pharmaceutical industries is of prime importance.

References

Abbadi A, Domergue F, Bauer J, Napier JA, Welti R et al (2004) Biosynthesis of very-long-chain polyunsaturated fatty acids in transgenic oilseeds: constraints on their accumulation. Plant Cell 16:2734–2748

Almasi K, Esnaashari SS, Khosravani M, Adabi M (2021) Yogurt fortified with omega-3 using nanoemulsion containing flaxseed oil: investigation of physicochemical properties. Food Sci Nutr 9(11):6186–6193

Almehmadi A, Lightowler H, Chohan M, Clegg ME (2021) The effect of a split portion of flaxseed on 24-h blood glucose response. Eur J Nutr 60:1363–1373

Álvarez MF, Mosquera T, Blair MW (2014) The use of association genetics approaches in plant breeding. In: Janick J (ed) Plant Breeding Reviews, Vol, vol 38, pp 17–68

Austria JA, Richard MN, Chahine MN (2008) Bioavailability of alpha linolenic acid in subjects after ingestion of three different forms of flaxseed. J Am Coll Nutr 27:214–221

Bernacchia R, Preti R, Vinci G (2014) Chemical composition and health benefits of flaxseed. Austin J Nutr Food Sci 2(8):1045

Caligiuri SP, Aukema HM, Ravandi A, Guzman R, Dibrov E, Pierce GN (2014) Flaxseed consumption reduces blood pressure in patients with hypertension by altering circulating oxylipins via an α-linolenic acid-induced inhibition of soluble epoxide hydrolase. Hypertension 64:53–59

Carter J (1993) Potential of flaxseed and flaxseed oil in baked goods and other products in human nutrition. Cereal Foods World 38(10):753–759

Chen Y, Zhou XR, Zhang ZJ, Dribnenki P, Singh S, Green A (2015) Development of high oleic oil crop platform in flax through RNAi-mediated multiple FAD2 gene silencing. Plant Cell Rep 34: 643–653

Chung MWY, Lei B, Li-Chan ECY (2005) Isolation and structural characterization of the major protein fraction from NorMan flaxseed (*Linum usitatissimum* L.). Food Chem 90:271–279

Cloutier S, Niu Z, Datla R, Duguid S (2009) Development and analysis of EST-SSRs for flax (*Linum usitatissimum* L.). Theor Appl Genet 119:53–63

Cloutier S, Ragupathy R, Miranda E, Radovanovic N, Reimer E et al (2012) Integrated consensus genetic and physical maps of flax (*Linum usitatissimum* L.). Theor Appl Genet 125:1783–1795
Czemplik M, Kulma A, Bazela K, Szopa J (2012) The biomedical potential of genetically modified flax seeds overexpressing the glucosyltransferase gene. BMC Complement Altern Med 12:251
Das J, Naik B (2017) Linseed: a valuable crop plant. Int J Adv Res 5(3):1428–1442
Deng X, Long SH, He DF, Li X, Wang YF et al (2011) Isolation and characterization of polymorphic microsatellite markers from flax (*Linum usitatissimum* L.). Afr J Biotechnol 10 (5):734–739
Dhirhi N, Mehta N, Singh S (2018) Estimation of heterosis for seed yield and its attributing traits in linseed (*Linum usitatissimum* L.). Int J Curr Microbiol Appl Sci 7(11):2332–2341
Diederichsen A, Raney JP, Duguid SD (2006) Variation of mucilage in flax seed and its relationship with other seed characters. Crop Sci 46:365–371
Dmitriev AA, Kezimana P, Rozhmina TA, Zhuchenko AA, Povkhova LV et al (2020) Genetic diversity of SAD and FAD genes responsible for the fatty acid composition in flax cultivars and lines. BMC Plant Biol 20:301. https://doi.org/10.1186/s12870-020-02499-w
Dutta S, Mukherjee PK, Mitra A, Zaman S, Mandal AB (2021) Efficient *in planta* genetic transformation through agrobacterium-mediated floral-dip technique in Indian flax. Plant Cell Biotechnol Mol Biol 22(53&54):21–36
Dyer D (2014) Flaxseeds and breast cancer. https://www.oncologynutrition.org/erfc/healthy-nutrition-now/foods/flaxseeds-and-breast-cancer
Eric-Parfait Kouamé KJ, Fanny A, Bora M, Li X, Sun Y, Liu L (2021) Novel trends and opportunities for microencapsulation of flaxseed oil in foods: a review. J Funct Foods 87:104812
Fahs Z, Rosseza Y, Guénin S, Gutierrez L, Thomasseta B, Perrin Y (2019) Cloning and molecular characterization of three lysophosphatidic acid acyltransferases expressed in flax seeds. Plant Sci 280:41–50
FAOSTAT (2020). https://www.fao.org/faostat/en/#data/QCL
Fu YB (2019) A molecular view of flax gene pool, chapter 2. In: Cullis CA (ed) Genetics and genomics of Linum, plant genetics and genomics: crops and models 23, https://doi.org/10.1007/978-3-030-23964-0_2
Fujisawa M, Watanabe M, Choi SK, Teramoto M, Ohyama K et al (2008) Enrichment of carotenoids in flaxseed (*Linum usitatissimum* L.) by metabolic engineering with introduction of bacterial phytoene synthase gene crtB. J Biosci Bioeng 105:636–641
Gaber DA, Badawy WA (2019) Role of flaxseed oil and silymarin in amelioration of lead-induced kidney injury. Kasr Al Ainy Med J 25:29–37
Ganorkar PM, Jain RK (2013) Flaxseed – a nutritional punch. Int Food Res J 20:519–525
Gepts P, Papa R (2003) Possible effects of (trans) gene flow from crops on the genetic diversity from landraces and wild relatives. Environ Biosaf Res 2:89–103
Gowda V, Sharma A, Goyal AK, Singh S, Arora (2018) Process optimization and oxidative stability of omega-3 ice cream fortified with flaxseed oil microcapsules. J Food Sci Technol 55(5):1705–1715
Green AG, Chen Y, Singh SP, Dribnenki (2008) Flax. In: Kole C, Hall TC (eds) Compendium of transgenic crop plants: transgenic oilseed crops. Blackwell Publishing, Chicester, UK, pp 200–226
Hall LM, Booker H, Siloto RMP, Jhala AJ, Weselake RJ (2016) Flax (*Linum usitatissimum* L.). Industrial Oil Crops. AOCS press, pp 157–194
Hano C, Martin I, Fliniaux O, Legrand B, Gutierrez L et al (2006) Pinoresinol-lariciresinol reductase gene expression and secoisolariciresinol diglucoside accumulation in developing flax (*Linum usitatissimum* L.) seeds. Planta 224:1291–1301
Harlan JR, de Wet JMJ (1971) Towards a rational classification of cultivated plants. Taxon 20:509–517
He L, Xiao J, Rashid KY, Yao Z, Li P et al (2019) Genome-wide association studies for pasmo resistance in fax (*Linum usitatissimum* L.). Front. Plant Sci 9:1982
Hoque A, Fiedler JD, Rahman M (2020) Genetic diversity analysis of a flax (*Linum usitatissimum* L.) global collection. BMC Genomics 21:557

https://www.indexbox.io/blog/which-country-consumes-the-most-linseed-oil-in-the-world/
Ibrahim AK, Zhang L, Niyitanga S, Afzal MZ, Xu Y et al (2020) Principles and approaches of association mapping in plant breeding. Trop Plant Biol 13:212–224
Jagtap AA, Badhe YS, Hegde MV, Zanwar AA (2020) Development and characterization of stabilized omega-3 fatty acid and micronutrient emulsion formulation for food fortification. J Food Sci Technol 58(3):996–1004
Kesiraju K, Tyagi S, Mukherjee S, Rai R, Singh NK et al (2021) An apical meristem-targeted *in planta* transformation method for the development of transgenics in flax (*Linum usitatissimum*): optimization and validation. Front Plant Sci 11:562056. https://doi.org/10.3389/fpls.2020.562056. eCollection 2020
Korbes AP, Kulcheski FR, Margis R, Margis-Pinheiro M, TurchettoZolet AC (2016) Molecular evolution of the lysophosphatidic acid acyltransferase (LPAAT) gene family. Mol Phylogenet Evol 96:55–69
Kumar N, Kumar V (2021) Assessment of genetic diversity in linseed germplasm using morphological traits. Electron J Plant Breed 12(1):66–73
Kumar S, You FM, Cloutier S (2012) Genome wide SNP discovery in flax through next generation sequencing of reduced representation libraries. BMC Genomics 13:684
Kumar S, You FM, Duguid S, Booker H, Rowland G et al (2015) QTL for fatty acid composition and yield in linseed (*Linum usitatissimum* L.). Theor Appl Genet 128(5):965–984
Lorenc-Kukuła K, Amarowicz R, Oszmiański J, Doermann P, Starzycki M et al (2005) Pleiotropic effect of phenolic compounds content increases in transgenic flax plant. J Agric Food Chem 53:3685–3692
Ludvíková M, Griga M (2015) Transgenic flax/linseed (*Linum usitatissimum* L.) – expectations and reality. Czech J Genet Plant Breed 51(4):123–141
Luo Z, Iaffaldano BJ, Zhuang X, Fresnedo-Ramirez J, Cornish K (2017) Analysis of the first Taraxacum kok-saghyz transcriptome reveals potential rubber yield related SNPs. Sci Rep 7(1):9939
Marambe P, Shand P, Wanasundara P (2008) An *in-vitro* investigation of selected biological activities of hydrolysed flaxseed (*Linum usitatissimum* L.) proteins. J Am Oil Chem Soc 85:1155–1164
McKenzie RR (2011) Genetic and hormonal regulation of stem vascular tissue development in flax (*Linum usitatissimum* L.). MSc Thesis, University of Alberta, Edmonton, Alberta, pp 1–220
Meacham J, Ajmera R, Tan V (2022) Top 9 Health benefits of flax seeds. https://www.healthline.com/nutrition/benefits-of-flaxseeds#TOC_TITLE_HDR_1
Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genomewide dense marker maps. Genetics 157:1819–1829
Mohammadi AA, Saeidi G, Arzani A (2010) Genetic analysis of some agronomic traits in flax (*Linum usitatissimum* L.). Aust J Crop Sci 4:343–352
Murugkar DA, Zanwar A, Shrivastava A (2021) Effect of nano-encapsulation of flaxseed oil on the stability, characterization and incorporation on the quality of eggless cake. Appl Food Res 1(2):100025
Nag S, Mitra J, Karmakar PG (2015) An overview on flax (*Linum usitatissimum* L.) and its genetic diversity. Int J Agric Environ Biotechnol 8(4):805–817
Nowak DA, Snyder DC, Brown AJ, Wahnefried WD (2007) The effect of flaxseed supplementation on hormonal levels associated with polycystic ovarian syndrome: a case study. Curr Top Nutraceutical Res 5:177–181
Ntiamoah C, Rowland GG (1997) Inheritance and characterization of two low linolenic acid EMS-induced McGregor flax (*Linum usitatissimum*). Can J Plant Sci 77:353–358
Oh TJ, Gorman M, Cullis CA (2000) RFLP and RAPD mapping in flax (*Linum usitatissimum*). Theor Appl Genet 101:590–593
Omoni AO, Aluko RE (2006) Mechanism of the inhibition of calmodulin-dependent neuronal nitric oxide synthase by flaxseed protein hydrolysates. J Am Oil Chem Soc 83:335–340

Panse ML, Phalke SD (2016) Fortification of food with Omega-3 fatty acids. In: Hegde MV, Zanwar AA, Adekar SP (eds) Omega-3 fatty acids - keys to nutritional health. Springer International Publishing, Switzerland, pp 89–100

Patel A, Desai SS, Mane VK, Enman J, Rova U et al (2022) Futuristic food fortification with a balanced ratio of dietary ω-3/ω-6 omega fatty acids for the prevention of lifestyle diseases. Trends Food Sci Technol 120:140–153

Patial R, Paul S, Sharma D, Sood VK, Kumar N (2019) Morphological characterization and genetic diversity of linseed (*Linum usitatissimum* L.). J Oilseeds Res 36(1):8–16

PGRC (Plant Gene Resource of Canada) (2018) Entries for Taxa in the GRIN-CA database. http://pgrc3.agr.gc.ca/cgi-bin/npgs/html/tax_search_new.pl

Prabha S, Yadav A, Yadav HK, Kumar S, Kumar R (2017) Importance of molecular markers in linseed (*Linum usitatissimum*) genome analysis–a review. Crop Res 52(1, 2 & 3):61–66

Radovanovic N, Thambugala D, Duguid S, Loewen E, Cloutier S (2014) Functional characterization of flax fatty acid desaturase FAD2 and FAD3 isoforms expressed in yeast reveals a broad diversity in activity. Mol Biotechnol 56:609–620

Ragupathy R, Rathinavelu R, Cloutier S (2011) Physical mapping and BAC-end sequence analysis provide initial insights into the flax (*Linum usitatissimum* L.) genome. BMC Genomics 12:217

Ratnayake WMN, Galli C (2009) Fat and fatty acid terminology, methods of analysis and fat digestion and metabolism: a background review paper. Ann Nutr Metab 55:8–43

Renouard S, Tribalatc MA, Lamblin F, Mongelard G, Fliniaux O et al (2014) RNAi-mediated pinoresinol lariciresinol reductase gene silencing in flax (*Linum usitatissimum* L.) seed coat: consequences on lignans and neolignans accumulation. J Plant Physiol 171:1372–1377

Rezaei S, Sasani M, Akhlaghi M, Kohanmoo A (2020) Flaxseed oil in the context of a weight loss programme ameliorates fatty liver grade in patients with non-alcoholic fatty liver disease: a randomised double-blind controlled trial. Br J Nutr 123(9):994–1002

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F et al (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet 44:217–220

Rowland GG (1991) An EMS-induced low-linolenic-acid mutant in McGregor flax (*Linum usitatissimum* L.). Can J Plant Sci 71:393–396

Rui-López N, Haslam RP, Venegas-Calerón M, Larson TR, Graham IA et al (2009) The synthesis and accumulation of stearidonic acid in transgenic plants: a novel source of heart-healthy omega-3 fatty acids. Plant Biotechnol J 7:04–716

Saivarshine S, Kavitha S, Vishnupriya V, Gayathri R (2020) *In vitro* xanthine oxidase inhibitory potential of flaxseed oil. Plant Cell Biotechnol Mol Biol 21(29–30):92–96

Singh KK, Mridula D, Rehal J, Barnwal P (2011) Flaxseed: a potential source of food, feed and fiber. Crit Rev Food Sci Nutr 51:210–222

Soto-Cerda BJ, Butler IM, Muñoz G, Rupayan A, Cloutier S (2012) SSR-based population structure, molecular diversity and linkage disequilibrium analysis of a collection of flax (*Linum usitatissimum* L.) varying for mucilage seed coat content. Mol Breed 30(2):875–888

Soto-Cerda BJ, Duguid S, Booker H, Rowland G, Diederichsen A et al (2014) Association mapping of seed quality traits using the Canadian flax (*Linum usitatissimum* L.) core collection. Theor Appl Genet 127:881–896

Soto-Cerda BJ, Cloutier S, Quian R, Gajardo HA, Olivos M et al (2018) Genome-wide association analysis of mucilage and hull content in flax (*Linum usitatissimum* L.) seeds. Int J Mol Sci 19:2870

Speck A, Trouve JP, Enjalbert J, Geffroy V, Joets J et al (2022) Genetic architecture of powdery mildew resistance revealed by a genome-wide association study of a worldwide collection of flax (*Linum usitatissimum* L.). Front Plant Sci 13:871633

Stefani FDS, Campo CD, Paese K, Guterres SS, Costa TMH et al (2019) Nanoencapsulation of linseed oil with chia mucilage as structuring material: characterization, stability and enrichment of orange juice. Food Res Int 120:872–879

Stuglin C, Prasad K (2005) Effect of flaxseed consumption on blood pressure, serum lipids, hemopoietic system and liver and kidney enzymes in healthy humans. J Cardiovasc Pharmacol Ther 10(1):23–27

Terfa GN, Gurmu GN (2020) Genetic variability, heritability and genetic advance in linseed (*Linum usitatissimum* L.) genotypes for seed yield and other agronomic traits. Oil Crop Sci 5:156–160

Vavilov NI (1935) Theoretical basis for plant breeding, Moscow. Origin and geography of cultivated plants. In: Love D (trans.) The phytogeographical basis for plant breeding, vol 1. Cambridge University Press, Cambridge, UK, pp 316–366

Vrinten P, Hu Z, Munchinsky MA, Rowland G, Qiu X (2005) Two FAD3 desaturase genes control the level of linolenic acid in flax seed. Plant Physiol 139:79–87

Wang WJ, Qiu C, Ye Y, Guo X et al (2017) Comparison of phytochemical profiles and health benefits in fiber and oil flaxseeds (*Linum usitatissimum* L.). Food Chem 214:227–233

Wu J, Zhao Q, Wu G, Zhang S, Jiang T (2017) Development of novel SSR markers for flax (*Linum usitatissimum* L.) using reduced-representation genome sequencing. Front Plant Sci. https://doi.org/10.3389/fpls.2016.02018

Xie D, Dai Z, Yang Z, Tang Q, Sun J et al (2018) Genomic variations and association study of agronomic traits in flax. BMC Genomics 19:512

Yi L, Gao F, Siqin B, Zhou Y, Li Q et al (2017) Construction of an SNP-based high-density linkage map for flax (*Linum usitatissimum* L.) using specific length amplified fragment sequencing (SLAF-seq) technology. PLoS One 12(12):e0189785. https://doi.org/10.1371/journal.pone.0189785

You FM, Zheng C, Bartaula S, Khan N, Wang J, Cloutier S (2022) Genomic cross prediction for linseed improvement. In: Gosal SS, Wani SH (eds) Accelerated plant breeding. Spinger Nature, Cham, pp 451–480

Zhang J, Qi Y, Wang L, Wang L, Yan X et al (2020) Genomic comparison and population diversity analysis provide insights into the domestication and improvement of flax. iScience 23:100967

Zhang J, Xie Y, Miao C, Wang L, Zhao W et al (2022) Secoisolariciresinol diglycoside (SDG) lignan content of oil flax: genotypic and environmental variations and association with other traits. Oil Crop Sci 7(1):1–8

Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR et al (2016) Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet 48:481

Zuk M, Prescha A, Stryczewska M, Szopa J (2012) Engineering flax plants to increase their antioxidant capacity and improve oil composition and stability. J Agric Food Chem 60:5003–5012

Increasing Nutraceutical and Pharmaceutical Applications of Safflower: Genetic and Genomic Approaches

Megha Sharma, Varun Bhardwaj, Poulami Goswami, Anmol Kalra, Kadirvel Palchamy, Arun Jagannath, and Shailendra Goel

Contents

Abstract

Nutraceuticals are functional foods that are produced in nature and show medicinal, health-promoting, and disease-preventing properties. One of the important sources of dietary and herb-based nutraceuticals is safflower. Safflower is an oil seed crop cultivated globally in ~23 countries. Being a source of unsaturated fatty acids such as linoleic and oleic acid, safflower oil confers several health benefits. Besides seed oil, petals and leaves of the safflower have been used widely in the pharmaceutical industry due to their richness in flavonoids and alkaloids. Crude floral extracts are known to possess anti-inflammatory and analgesic properties and bring wide health-related benefits in cardiovascular and diabetic patients. Although safflower oil is healthier, it is not a popular crop due to its lower oil content and economical value. Various efforts have been made by plant scientists

Megha Sharma, Varun Bhardwaj and Poulami Goswami contributed equally to this study.

M. Sharma · V. Bhardwaj · P. Goswami · A. Kalra · A. Jagannath (✉) · S. Goel (✉)
Department of Botany, University of Delhi, Delhi, India

K. Palchamy
ICAR-Indian Institute of Oilseeds Research, Hyderabad, Telangana, India

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_20

to increase oil content as well as oil composition in the safflower. Over the past few decades, various breeding strategies including conventional breeding, molecular breeding, and genetic engineering approaches resulted in improved cultivars with better oil composition and content. Thus, with its high nutrition value, safflower has great potential to play an important role in the nutraceutical and pharmaceutical industry.

Keywords

Safflower · Nutraceutical · Pharmaceutical · Seed-based-product · Non-seed based product · Alkaloids · Flavonoids

1 Introduction

With the emergence of the new world and increased life expectancy, the risk of lifestyle diseases has also increased significantly. Epidemiological studies show that dietary habits play a critical role in human health. These studies increased the popularity of plant-based compounds commonly called "nutraceuticals." The combined efforts of the food industry, pharmaceutical, and newly merged nutritional conglomerates led to the development of nutraceuticals to cater to demands of the herbal and dietary supplements market. Nutraceuticals are often referred as functional foods or phytochemicals. These are natural, bioactive, chemical compounds with medicinal, health-promoting, and disease-preventing properties (Dureja et al. 2003). They can be classified as Nutrients, Herbal, or Dietary based compounds. Herbal or botanical products are concentrated or extracts whereas Dietary products are the reagents derived from other sources which include pyruvate, sulphate. These products serve functions of sports nutrients and meal replacement. In the present era, Aloe Vera, Echinacea, Evening primrose oil, Garlic, and Ginger Plantago have been used commonly in herbal and phytochemical products (Dureja et al. 2003).

The human staple diet includes fruits, vegetables, whole grains, herbs, nuts, and seeds and is often supplemented with fats and edible oil which are rich in energy. Based on the saturation levels, edible oil rich in unsaturated fatty acids are healthier as compared to oils rich in saturated fatty acids. Safflower oil is characterized by higher levels of unsaturated fatty acids, and the plant is also a rich source of secondary metabolites. Hence, safflower has immense potential as a nutraceutical crop and can be an important source of herbal as well as dietary supplements (Fig. 1).

Safflower (*Carthamus tinctorius* L.*)* is a member of the Asteraceae. It is grown in ~23 countries with 6.2 million tons of seed production (FAOSTAT 2019). Kazakhstan, United States of America (USA), Russia, Mexico, China, and India are the major producers of safflower and contribute to ~81% of the global production. Along with oil, it has been used as coloring dye, food flavoring agent, and for pharmaceutical purposes (Carlsson et al. 2014). More than 200 compounds can be extracted from safflower including fatty acids, flavonoids, and steroids.

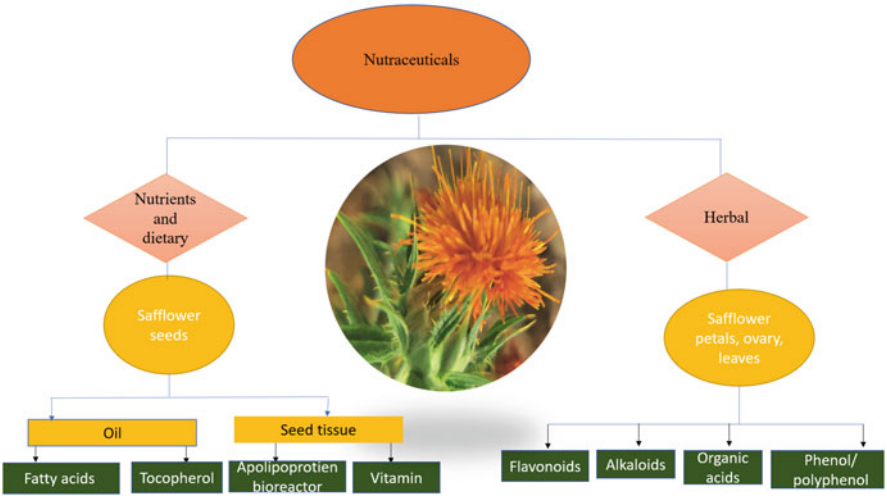


Fig. 1 Scope of nutraceuticals in safflower *(Carthamus tinctorius)*

Safflower promises diverse roles in human nutrition and health. The nutritionally beneficial compounds are mainly extracted from its seeds and flowers. Based on the plant part used, safflower products can be categorized into seed-based and non-seed based products. Safflower seeds are the source of edible oil along with protein, β-carotene (vitamin A), and riboflavin (vitamin B_2). The richness of safflower oil in unsaturated fatty acids such as OA and LA along with low levels of low-density lipoprotein makes it highly effective in managing cardiovascular diseases. It contributes to artery dilation, lowering of blood pressure, blood flow enhancement as well as tissue oxygenation. Due to the high levels of LA, it is also effective in controlling osteoporosis. It possesses anti-inflammatory properties, corrects bone loss, and increases intestinal calcium absorption (Bae et al. 2002). The seed has active α-glucosidase inhibitors in derivative form of serotonin and is used in traditional medicine for treatment of diabetes (Takahashi and Miyazawa. 2012). The non-seed based plant parts such as petals are the source of the flavonoid glycosides, carthamin, kaempferol, quercetin, chalcones including hydroxysafflor yellow A (Shirwaikar et al. 2010). Safflower yellow A showed inhibition of formaldehyde-induced swelling, formation of ball granuloma, and histamine-stimulated capillary permeability and thus exhibits anti-inflammatory effects (Huang et al. 1984). Petals of safflower possess analgesic activity and can be considered a potential drug equivalent to morphine without the side effects (Almeida et al. 2001). Being rich in hypolipidemic agents, safflower petals have antidiabetic effects. The flowers of safflower elevate insulin level by enhancing insulin secretion from islets of Langerhans as well as regeneration, restores protein breakdown, and promotes glycogenesis (Asgary et al. 2012). Being the source of the fatty acids, tocopherols, vitamins, and antioxidants, safflower oil is a potential bioactive food ingredient.

Safflower exhibits a high level of geographical diversity (Ashri and Knowles 1960), phenotypic diversity (Ashri et al. 1974; as well as molecular diversity (Hassani et al. 2020; Ambreen et al. 2015; Amini et al. 2008). Safflower is characterized by a rich germplasm, consisting of cultivars, landraces, and wild-species. A germplasm directory of safflower documented details of germplasm collected from 18 different collection sites across 14 countries (Zhang and Johnson 1999) (http://safflower.wsu.edu/). Currently, a large collection of safflower germplasm is available at USDA, USA and has a repository of ~2400 accessions representing germplasm from 50 countries. A large collection of germplasm and the associated genetic diversity provide an opportunity to exploit safflower for its nutraceutical as well as pharmaceutical properties through genetic improvement programs. The focus of such breeding efforts has largely been on improvement of seed-based products while other non-seed based compounds are yet to be targeted in research programs. Thus, this review on safflower provides a detailed study of the composition, utility, and genetic improvement of seed-based products of safflower, along with a brief account on non-seed based products.

2 Seed-Based Compounds in Safflower

2.1 Description of Compounds

Safflower is well supplied with compounds such as fatty acids (including OA and LA), vitamins such as tocopherols, and antioxidants such as flavonoids and alkaloids (Table 1). Although flavonoids and alkaloids could be extracted from seeds, they are major constituents of petals and are described in later sections.

2.1.1 Fatty Acids and Oils

Oilseed crops are major sources of vegetable oil. Based on seed oil content and fatty acid composition, they can be utilized for dietary, industrial, or pharmacological purposes. Oil content in safflower germplasm ranges from 20% to 45%. The crop germplasm can be broadly divided into two categories (High Oleate and High Linoleate varieties) based on composition of seed oil, which is principally composed of two unsaturated fatty acids: Oleic acid (C18:1) (OA) and Linoleic acid (C18:2) (LA). These account for approximately 90% of total fatty acid content in safflower oil. The remaining 10% is contributed by saturated fatty acids such as Palmitate (6–8%) and Stearate (2–3%), with traces of other fatty acids not exceeding 0.81% (Porokhovinova et al. 2022). In a study conducted by Fernández-Martinez et al. (1993), fatty acid compositions of 200 safflower accessions were analyzed and both OA and LA demonstrated significant levels of variation, from 3.1% to 90.60% and from 3.9% to 88.8%, respectively.

Safflower seed oil has also been used for treating cardiac and hepatic disorders (Dajue and Mündel 1996). Recent reports have suggested that, when other vegetable oils were blended with safflower oil it enhanced their nutritional value/quality. It has been shown to improve various physiological systems and is used for treatment of

Table 1 Compounds isolated and characterized from safflower seeds (Adapted and modified from Zhang et al. 2016)

S. No	Classification of compound	Name of compound	Reference
1.	Organic acids	Oleic acid	
		Linoleic acid	
		Palmitic acid	Gecgel et al. (2007), Fernandez-Martinez et al. (1993)
		Stearic acid	
		Myristic acid	
		Palmitoleic acid	Ghareghani et al. (2017)
		Linolenic acid	
		Caffeic acid	Vieira et al. (2021)
		Ferulic acid	Zhang et al. (1997)
2	Flavonoids	Kaempferol 7-O-β-D-glucopyranoside	
		Acacetin–glucuronide pentoside	Chakradhari et al. (2020)
		Acacetin-7-O-D-glucuronide	
3	Alkaloids	N-feruloyl serotonin	Koyama et al. (2006)
		N-(p-coumaroyl) serotonin	Takii et al. (1999)
		N-(p-coumaroyl)serotonin-O-β-D-glucopyranoside	Sakamura et al. (1978)
		N-[2-(5-hydroxy-1H-indol-3-yl)ethyl] ferulamide	Zhang et al. (1997)
4	Other compounds	Coniferyl alcohol	Peng et al. (2017)
		Sinapyl alcohol	Peng et al. (2017)

arteriosclerosis, hyperlipidaemia, and coronary heart disease (Abidi 2001). The polyunsaturated fatty acids (PUFA) in safflower oil suppress the accumulation of low-density lipoproteins (LDL or bad cholesterol) (Cho 2001). Consumption of safflower phospholipids was shown to have desirable effects on human health such as reduction of lipids in liver and higher levels of high density lipoprotein (Iwata et al. 1992). Significant reduction was observed in levels of total cholesterol, blood glucose, triglyceride LDL, alkaline phosphatase, alanine aminotransferase, and aspartate aminotransferase in alloxan-induced diabetic rats after their diet was supplemented with 200 mg/kg safflower seed oil, demonstrating the protective effects of safflower oil on liver (Rahimi et al. 2014). In the same study, it was concluded that safflower oil had hypoglycemic and hypolipidemic effect in hyperglycemic rats under the same set of dietary conditions, adding to the growing evidence on its antidiabetic and anti-obesity properties. Zhang et al. (2010) showed that diet supplemented with safflower oil can change gene expression pertaining to fat deposition resulting in mitigation of diet-induced obesity. Safflower oil can stabilize storage of lipophilic compounds showing potential as a functional food or for usage as vector in drug delivery. Cod liver oil encapsulated in safflower oil has increased oxidative stability of PUFAs.

2.1.2 Tocopherols

Tocopherols ($C_{28}H_{48}O_2$), commonly known as vitamin E, is a group of closely related lipids with substitutions on 2H-1-benzopyran-6-ol nucleus along with isoprenoid units. The oilseeds have different types of tocopherols, viz. α-, β-, γ-and δ-tocopherols which protect PUFA against oxidation (Vosoughkia et al. 2011). Three types of tocopherols (α-tocopherol, β-tocopherol, and γ-tocopherol) were found in safflower oil in various amounts ranging from 460.5 to 709.3 mg/kg, 8.5 to 21.6 mg/kg, and 4.5 mg/kg of oil, respectively (Mani et al. 2020). In a separate study, cold pressed safflower seed oil was found to contain α-tocopherol at 376.6 mg/kg, as the main component along with β-tocopherol (4.4 mg/kg), γ-tocopherol (25.9 mg/kg), and δ-tocopherol (1.0 mg/kg) (Topkafa 2016). Safflower oil has high α-tocopherol content making it an excellent source of vitamin E, although low thermostability makes it less suitable for deep frying (Cuesta et al. 2014). Tocopherols have also been used for feed, food, resins, pharmaceuticals, and cosmetics.

2.2 Genetic Improvement Efforts in Safflower

With an ever-increasing demand for healthier foods and plant-based nutraceutical products, research groups around the world are exploring the natural diversity of compounds found in safflower. This entails elucidation of molecular mechanisms governing the biosynthetic and regulatory pathways to modify the biochemical machinery for production of desired products. With continuous efforts in safflower breeding, nutraceutical profile of safflower seeds has been enhanced, especially for oleic acid. However, to keep pace with current market demands further improvement along the same lines is required. Evaluation of safflower germplasm and identification of lines with differing fatty acid profiles could serve as basis for development of lines with higher oil and tocopherol content. Modifying amino acid composition of seed meal as well as purging of harmful content like matairesinol monoglucoside and lignin glucoside can further enhance the nutritional quality of safflower seeds.

2.2.1 Conventional Breeding

The initiation of breeding programmes for crop improvement require germplasm possessing high genetic variability for quantitative traits. Many institutions across the world have curated germplasm collections from major centres of safflower diversity including national collections at China, India, and USA. Detailed information about safflower germplasm collections can be found in a review by Vollman and Istvan (2010).

Oil content and composition are complex traits and are significantly influenced by genotype X environment interactions. Therefore, breeders have relied on simple traits with high heritability for identification of superior genotypes. The direct attributes for oil content and oil composition are total yield of the plant, whereas the indirect attributes for seed yield are head (capitulum) number, head diameter, and grains per head (Arslan 2007). For oil composition, head weight, number of heads, and hull (pericarp) content was used for indirect selection. Seed structure is also an

important determinant of oil content. The percentage of oil in seed shows inverse correlation with hull size. The percentage of oil increases in the very thin hull types (Bergman et al. 1985). Other factors which affect the yield and oil content are crop phenology which is related to genotype and environmental condition.

Biometric analyses have been performed to estimate the effect of genes on morphological traits. Number of branches and plant height are under the influence of additive gene action whereas head diameter (Golkar et al. 2012) shows low broad-sense heritability. The white hull is dominant over the striped hull (Urie 1986). The striped seed and reduced pericarp were shown to be controlled by recessive genes (*Th* and *Stp*) (Ebert and Knowles 1966) and had monogenic inheritance (Ashri and Efron 1964. The seed protein content is governed by additive dominance model (Golkar et al. 2012). Besides oil content, major attention has been given to oil composition for safflower improvement as it makes safflower oil a healthier alternative to others plant-based oils available in the market. Safflower oil is rich in the OA, LA, Palmatic acid (PA) and Stearic acid (SA). For fatty acid composition, both narrow and broad-sense heritabilities have been reported (Golkar et al. 2011). All the fatty acids follow additive gene effect [OA (Hamdan et al. 2009b), LA (Hamdan et al. 2008), PA & SA (Hamdan et al. 2009a)]. LA and SA in safflower are influenced by maternal factors (Golkar et al. 2011) whereas high OA is under the influence of recessive alleles (Fernandez–Martinez et al. 1993).

Unlike fatty acids, safflower germplasm shows little variability for tocopherol profiles. Johnson et al. 1999 reported no variability in the tocopherol profiles whereas Velasco and Fernández-Martínez (2001) found variation in γ-tocopherol profiles. Velasco et al. (2005) identified a natural mutant of *C. oxycantha* which is rich in high gamma-tocopherol instead of standard high alpha-tocopherol content. Mutant allele from *C. oxycantha* was then introgressed in C. tinctorius followed by selection, resulting in a line with 90% gamma-tocopherol content (Velasco et al. 2005).

Conventional breeding methods, however, have had limited success in improving quantitatively inherited characters listed above. Crossing and record keeping procedures are laborious and slow, compounded with limitations such as low genetic base in varieties, loss of genes in the successive generations due to segregation and genetic drag. These limitations can be mitigated to some extent by applying recent genomic tools.

2.2.2 Molecular Breeding

Modern tools such as expression profiling and genetic engineering have been effectively used to address limitations of classical breeding efforts. Linkage maps are used to identify chromosomal regions that control qualitative and quantitative traits (Collard et al. 2005); DNA markers associated with genes are subsequently used for marker-assisted selection (MAS). Breeding objectives so far have been primarily focused on increasing seed yield and oil yield and acquisition of resistance against biotic and abiotic stresses. Little has been done on the positional cloning of genes/QTLs involved in the production of nutraceutical compounds. Information on trait mapping in safflower is limited and is described below.

Construction of Linkage Maps

Linkage maps for safflower were initially constructed by Mayerhofer et al. (2010), using RFLP and PCR-based (SSR and cDNA) markers on an intraspecific F2 mapping population of *C. tinctorius* and an interspecific BC population generated through a cross between *C. tinctorius* X *C. oxycantha.* The resulting linkage maps were compared for synteny and showed significant collinearity of markers in several regions. Since then, several groups have attempted the construction of linkage maps (Table 2).

Linkage maps constructed by various groups in safflower over the last two decades suffer low resolution (with a few exceptions) as the number of markers utilized is limited. Mapping of traits in most cases was done using F2 populations which suffer from greater heterozygosity compared to RILs. It is therefore crucial that efforts be made for development of high-density linkage maps using SNP markers.

Mapping of Genes

Linoleic Acid

Inheritance of high LA content in safflower was first reported by Hamdan et al. (2008). Results from their study showed 5 RAPD bands linked to the LA content controlling *Li* locus. The fragments were sequenced to design SCAR markers and a linkage map was constructed to include 5 SCAR markers along with *Li* and *Ms* gene loci. The SCAR markers flanked *Li* and *Ms* loci at 15.7 cM and 3.7 cM respectively.

Table 2 Linkage maps constructed in safflower

References	Marker number and type	Population	No. of linkage groups	Length of map (cM)
Mayerhofer et al. (2010)	1412 SSR and 75 RFLP	F_2: *Carthamus tinctorius*	11	954
Mayerhofer et al. (2010)	1412 SSR and 75 RFLP	BC1: *C. tinctorius* X *C.oxycantha*	13	580
Hamdan et al. (2012)	47 RAPD, 60 SSR, and 4 SCAR	F_2: *Carthamus tinctorius*	15	816
Pearl et al. (2015)	244 SNP	F_2: *C. tinctorius X C. palestinius*	12	858
Karimi (2015)	168 SSR	F_2: *C. tinctorius*	11	877
Mirzahashemi et al. (2015)	119 SSR and ISSR	F_2: *C. tinctorius*	24	646
Bowers et al. (2016)	2008196 SNPs	F_6: RIL: *C. tinctorius* X C. *palestinius*	12	959
Wu et al. (2021)	248 SSR	F_2: AH04 X YH04 (both *C. tinctorius*)	12	1136.46

Adapted and modified from Golkar and Karimi (2019)

Oleic Acid

Most safflower germplasm lines are rich in linoleic acid (~70% of total oil content). Knowles (1989) identified safflower germplasm that produced high levels of oleic acid. When compared to oils with higher levels of polyunsaturation, oils high in oleic acid have been reported to have a hypocholesterolemic effect on human health and a higher oxidative stability (Mensink and Katan 1989). High oleic acid content is under the genetic control of partially recessive alleles at the *Ol* locus (Fernandez-Martinez et al. 1993). Efforts have been made for mapping major and modifying genes responsible for high oleic acid content (Hamdan et al. 2012). A genetic linkage map was constructed spanning 816.4 cM, comprising 15 linkage groups by using RAPD (47), SSR (60), and SCAR (4) marker loci using an F_2 population. This population was generated through a cross consisting of CL-1(male sterile) × CR-9 (>84% oleic acid) (Hamdan et al. 2012). The *Ol* locus was located on linkage group T3 between the SSR marker ct365 and the SCAR marker IASCA-73, respectively, with genetic distances of 0.4 and 39.1 cM.

Tocopherols

Tocopherols are the principal naturally occurring antioxidants in oilseed crops with protective action in biological systems and in oils or products derived from oils. Molecular tagging was used to decipher the genetic control of high gamma-tocopherol, leading to identification of markers for high gamma-tocopherol. A mutant line IASC-1 with high γ-tocopherol was crossed with CL-1 with high α-tocopherol. The resultant F2 population was analyzed for segregation of the partially recessive *Tph2* gene. Bulked segregant analysis revealed that *Tph2* gene was linked to eight RAPD and one SSR marker, leading to construction of a *Tph2* linkage map. RAPD fragments closest to the *Tph2* gene were converted into SCAR markers. Later, a γ-tocopherol methyltransferase locus was shown to co-segregate with *Tph2* (Garcia-Moreno et al. 2011).

Association Mapping in Safflower

Conventional methods are the time-consuming method for discovering genomic regions governing simple and/or complex traits, such as linkage analysis and/or QTL mapping, involves establishment of biparental mapping populations. Since fewer recombination events are evaluated in biparental population, the allelic variation obtained in QTL mapping is limited, resulting in poor mapping resolution. However, association mapping (AM) provides a faster and efficient method for evaluating complex traits at high resolution, offering a promising way to overcome the constraints of linkage mapping (Abdurakhmonov and Abdukarimov 2008). AM uses naturally occurring recombination events to identify correlations between phenotypes and genetic polymorphisms in a heterogeneous collection of unrelated accessions/genotypes, allowing for fine-scale mapping of attributes. It has proven to be an effective method for discovering marker-trait relationships for a variety of agronomic traits in several crop species (Zhang et al. 2014; Abdurakhmonov and Abdukarimov 2008). Small, operational core collections of safflowers have been constructed to assist genetic deconstruction of complex characteristics (Kumar et al. 2016). These core

collections can be the basis for AM if they represent substantial genetic variation and a weak population structure (Pritchard et al. 2000).

Ambreen et al. (2018) performed AM on a set of 124 safflower accessions (CartAP), resourced collectively from two core collections. Significant correlation ($R^2 > 10\%$, $P < 0.05$) was seen between 96 marker-traits associations (MTA). Consistent associations were represented in both General Linear and Mixed Linear models among two growing seasons for traits such as oil content, oil composition, and number of primary branches. Many MTAs were also found between parameters with positive or negative phenotypic correlations (e.g., plant height; days to 50% flowering; OA/LA concentration; number of primary branches/numbers of capitula per plant). Such associations will complement marker-assisted breeding and will assist in deciphering genetic basis of trait variability.

2.2.3 Marker Assisted Breeding and Gene Introgression

In the twenty-first century, molecular breeding has emerged as an indispensable tool for crop improvement crucial for meeting the continuous demand to increase the production of the crop A PCR based multiplex marker assay for selection of high OA allele *'ol'* based on mutation in *CtFAD2-1* gene was reported by Liu et al. (2013). Kadirvel et al. (2020) designed and used SNP genotyping assays such as KASP (Kompetitive Allele Specific PCR) and the Amplifluor™ SNPs genotyping systems for prediction of *"ol"* allele. Using these assays, the "*ol*" allele from Montola-2000 was introgressed into the Bhima using assisted backcrossing resulting in the development of promising high oleic lines with OA content ranging from 75.2% to 81.8%.

2.2.4 Genetic Engineering

Genetic engineering uses biotechnological tools to manipulate DNA, resulting in the transfer of genes within/between organisms to produce improved or novel organisms. When applied to crop improvement, transgenesis has yielded genetically modified (GM) crops having novel genes and desirable characteristics. It also overcomes the limitations of reproductive incompatibility encountered in conventional breeding methods. This allows for transfer of traits from any organism into the target crop. The following section describes studies in safflower pertaining to genetic engineering for improvement of health-related traits.

The nutraceutical, γ-linolenic acid (GLA) is a precursor for long-chain polyunsaturated fatty acids. It is synthesized from linoleic acid in endoplasmic reticulum by the activity of Δ-6 desaturase. Δ6-desaturase gene from *Saprleginadeclina* was transformed into a high LA variety of safflower resulting in accumulation of high levels of GLA (70% v/v). The transgenic line was eventually granted approval from US Food and Drug Administration and commercialized for nutraceutical purposes as Sonova 400 (Nykiforuk et al. 2012).

Safflower was also used as a bioreactor to produce Italian version of Apolipoprotein AI called $ApoAI_{Milano}$, which is a major component of circulating HDL and has therapeutic applications against atherosclerosis. In Atherosclerosis, plaque accumulates along the inner lining of an artery, causing hemorrhage or thrombosis (blood clot formation). $ApoAI_{Milano}$ was expressed in transgenic safflower as a fusion

protein specifically targeted to seed tissue. Results showed that selected lines accumulated 14.5% fusion protein as a fraction of total seed protein (TSP) (Nykiforuk et al. 2011). Similarly, recombinant human fibroblast growth factor 10 (rhFGF10) was expressed in safflower seeds using oil body-oleosin technology (Huang et al. 2017). Oleosin-rhFGF10 was introduced and expressed in safflower seed. The trait inheritance was checked till T3 generation.

Decades of plant breeding efforts have raised the oleic content in safflower to ~90%, but costed in terms of low field performance accompanied with poor yields. Wood et al. (2018) engineered super high oleic (SHO) safflower producing seed oil with 93% oleic acid. Safflower was transformed with hairpin-based vectors. Transgenic plants produced seed oil with low PUFA (<1.5%) and only 4% saturate fatty acids. There was no effect on lipid composition of leaves and roots. The seed yield of transgenics was also comparable to non-GM safflower even under different environments and varied sites.

2.2.5 Genomics-Aided Breeding

Genomics is a branch of biology primarily concerned with structure and function of genomes. Functional genomics, a subset of genomics, is concerned primarily with assigning functions to unknown genes. Various technological platforms such as transcriptomics, proteomics, metabolomics, and phenomics are used to analyze gene function. Numerous transcriptomic studies have been conducted to improve our knowledge and understanding of the molecular functions of the genomic components of various organisms. Such studies are expected to increase with the availability of genome sequences.

Partial cDNA clones were generated based on two protein sequences with thioesterase activity (molecular masses: 34 and 40 KDa). Similarly, cDNA clone (CTOS1) was generated based on a new protein putatively involved in accumulation of high levels of oleic acid in seeds (Mizukami et al. 2000) from an accession of safflower with high oleic acid in seeds. In recent years, sequencing of cDNA libraries using next generation sequencing (NGS) has emerged as a novel method for whole transcriptome analysis. De novo transcriptome analysis was performed by Li et al. (2012) using three different tissues, viz., leaves, petals, and seeds leading to identification of differentially expressed unigenes among these tissue. The analysis identified oleosins which were specifically expressed in the seeds. An enzyme, *FAD2* (microsomal oleoyl phosphatidylcholine desaturase) is known to introduce a double bond at Δ12 position of OA, thereby converting it to LA (Cao et al. 2013). They cloned 11 *ctFAD2* genes, each with a divergent functionality, the largest number of FAD2clones in any species. The temporal transcriptome profile of safflower for *FAD2* and other genes involved in oil accumulation was investigated by Li et al. (2021). Their study identified changes in expression of stearoyl-[acyl-carrier-protein] 9-desaturase gene (*SAD*) from 10 to 14 days after fertilization (DAF) and oleate desaturase (*FAD2–1*) from 14 to 18 DAF. They also identified 13 putative transcription factors (TFs) involved in regulating the expression levels of the *FAD2–1*gene.The study found a link between fatty acid biosynthesis and gene expression during seed development.

Safflower seeds (also known as *Carthami semen*) hold a wide array of nutritionally beneficial compounds with a greater part of compound diversity yet to be explored. The recent shift in consumer preferences towards healthier foods and plant-based pharma alternatives has reinvigorated research in improvement of nutritional aspects of seeds. Conventional breeding efforts have led to improvement in overall oil content while molecular breeding has led to the development of high oleic acid lines, many of which have been released for cultivation. With the publication of safflower genome and use of molecular breeding tools becoming commonplace, exploration into other classes of compounds such as seed flavonoids and alkaloids is to be expected.

3 Non-seed Based Compounds in Safflower

Safflower is cultivated primarily for its seeds and oil which could be exploited as nutraceuticals. However, since ancient times, safflower has been used in traditional medicines using/derived from flowers and leaves of the plant. Due to adverse effects of chemical medications, the pharmaceutical industry is exploring alternatives and Safflower has emerged as a promising option. This section describes pharmaceutical compounds obtained from safflower.

Flowers and leaves of safflower contribute medicinal as well as nutritional compounds. Safflower petals are consumed as herbal tea and have medicinal uses for menstrual problems, cardiovascular diseases, hypertension, arthritis, spondylosis, diabetes, pain, and swelling associated with trauma. They also have purgative, antioxidant, anti-inflammatory, analgesic, and anticonvulsant properties (Zhou et al. 2014). Flowers of the plant also possess central analgesic activity and can be considered as potential drugs equivalent to morphine-like substances. Thus, these can be beneficial to surpass the side effects caused by morphine (Almeida et al. 2001).

3.1 Description of Compounds

The major constituents of safflower flowers are flavonoids, glycosides, carthamin, and safflower yellow. Other components, viz., quercetin, carthamidin, kaempferol, isocarthamidin, glycosides of hydroxykaempferol, chalcones such as hydroxysafflor yellow A, safflomin-A, and acetylenic glucosides (carthamosides) have also been found (Table 3) (Jiang et al. 2005). The leaves of safflower are rich in Luteolin and its glucopyranosides (Lee et al. 2002).

3.1.1 Flavonoids

Flavonoids are polyphenolic secondary metabolites synthesized via phenylpropanoid pathway, with naringenin chalcone acting as precursor for a majority of them. In safflower, bioactive flavonoids are primarily concentrated in petals and seeds and exhibit a broad range of pharmacological activities. Flavonoids are broadly divided

Table 3 Compounds isolated and characterized from safflower flowers and leaves

S. No	Compound classification	Plant Part	Compound	Reference
1	Flavonoids	Flowers	Carthamin	Kanehira and Saito (1990)
			Safflor yellow A	Takahashi et al. (1982)
			Safflor yellow B	Takahashi et al. (1984)
			Safflomin A	Watanabe and Terabe (2000)
			Safflomin B	Chen et al. (2013)
			Tinctormine	Meselhy et al. (1993)
			Quercetin	Sun et al. (2003)
			Anhydrosafflor yellow B	Kazuma et al. (2000)
			Quercetin-3-O-β-D-glucoside	Kazuma et al. (2000)
			Quercetin-3-O-α-L-rhamnoside-7-O-β-D-glucuronide	Kazuma et al. (2000)
			Quercetin-7-O-β-D-glucoside	Hattori et al. (1992)
			Kaempferol	Kim et al. (1992)
			Kaempferol-3-O-β-D-glucoside	Hattori et al. (1992)
			Kaempferol-3-O-β-sophorose	Hattori et al. (1992)
			Kaempferol-3-O-β-rutinoside	Hattori et al. (1992)
			Scutellarein	Hattori et al. (1992)
		Leaves	51 Luteolin	
			52 Luteolin 7-O-β-D-glucopyranoside	Lee et al. (2002)
			53 Luteolin-7-0-(6″-O-acetyl)-β-D-glucopyranoside	
			Quercetin-7-O-(6″-O-acetyl)-β-D-glucopyranoside	
			Quercetin 7-O-β-D-glucopyranoside	
2	Alkaloids	Flowers	7,8-dimethyl pyrazino[2,3-g] quinazolin-2,4-(1H,3H) dione	Jiang (2008)
			Safflospermidine A	Jiang et al. (2008)
			Safflospermidine B	
3	Polyacetylenes	Flowers	4′,6′-acetonide-8Z-decaene-4,6-diyne-1-O-β-D-glucopyranoside	
			4,6-decadiyne-1-O-β-D-glucopyranoside	Zhou et al. (2006)
			(8Z)-decaene-4,6-diyne-1-O-β-D-glucopyranoside	

(continued)

Table 3 (continued)

S. No	Compound classification	Plant Part	Compound	Reference
			(8Z)-decaene-4,6-diyne-1-ol-1-O-β-D-glucuronyl-(1 0-20Þ-β-D-glucopyranoside	
			(2Z,8E)-tetradecadiene-4,6-diyne-1,12,14-triol-1-O-β-Dglucopyranoside	He et al. (2011)
			(2E,8Z)-tetradecadiene-4,6-diyne-1,12,14-triol-1-O-β-D-glucopyranoside	
4.	Organic acids	Flowers	p-coumaric acid	
			p-hydroxybenzoic acid	
			Succinic acid	Jiang (2008)
			4-O-β-D-glucopyranosyloxy-benzoic acid	
			4-O-β-D-glucosyl-trans-p-coumaric acid	Zhou et al. (2008)
			4-O-β-D-glucosyl-cis-p-coumaric acid	
5.	Other compounds	Flowers	Uridine	Jiang et al. (2008)
			Adenosine	
			Adenine	
			Thymine	
			Uracil	
			Roseoside	
			Sitosterol	
			Syringin	
			Methyl-3-(4-O-β-D-glucopyranosyl-3-methoxyphenyl) propionate	
			Ethyl-3-(4-O-β-D-glucopyranosyl-3-methoxyphenyl) propionate	Zhou et al. (2008)
			Ethylsyringin	
			Methylsyringin	

Adapted and modified from Zhang et al. (2016)

into two groups: quinochalcones and flavonols. The quinochalcones comprises of hydrosafflower yellow A, tinctorimine, carthamin, and cartorimine while the flavons include kaempferol, quercetin, and their glucosides (Guo et al. 2017).

Glycosides derived from flavonols such as quercetin and shannesol, have been extensively studied and were shown to possess antioxidative activity. Safflower extract containing flavonoids is also protective to the cardiac system, stabilizing

the oxygen supply to the heart and the heart rate (Guo et al. 2018). Flavonoids can induce aggregation of the platelets and depolymerization of ADP in platelets). The hypotensive effect of safflower flavones has been shown in animals (Mani et al. 2020). The major component of yellow pigment, Hydroxysafflor yellow A (HSYA) have a strong antagonistic effect on the receptors of platelet activating factor. In the in vitro experiments carried out by Zhang et al. 2002, varying concentrations of HSYA was found to inhibit both aggregation of polymorphonuclear leukocytes and platelet induced by platelet activating factor. HSYA also helps to improve the condition of acute myocardial inflammation (Zhou et al. 2013) by inhibiting the process causing it. HSYA also act as neuroprotector against cerebral injury and ischemia-reperfusion conditions by antioxidation, reducing neurological-deficit scores and decreasing superoxide-dismutase activity.

Diabetes is an important cause of vascular complications because it can activate dysfunctional biochemical pathways. One such example is the increased glycation of the proteins by methylglyoxal. Under in vitro conditions, HSYA was seen to inhibit protein glycation by reducing the production of advanced glycated end products, by reducing the protein modifications that were methylglyoxal-mediated and their cross-linking (Yue et al. 2013). The extract derived from the flowers of safflower also elevates insulin level by regenerating Langerhans islets and further induces the beta cells to secrete insulin. It is also capable of restoring breakdown of protein and promoting glycogenesis.

Quinochalcones are the primary yellow and red pigments in safflower. Structurally, most quinochalcones isolated from this plant have a unique C-glycosylated cyclohexanonedienol moiety. Up to 18 quinochalcone C-glycosides have been isolated from safflower till date. Carthamin is composed of two C-glucosylquinochalcone moieties (Zhang et al. 2016). Quinochalcone C-glycosides has an array of bioactivities like anticoagulation, anti-inflammation, antioxidation along with antihypertensive, and antitumor activity.

3.1.2 Alkaloids

Alkaloids are a group of small cyclic organic compounds, containing a nitrogen atom usually within the carbon ring, which gives them a slightly basic property. They are classified on the basis of precursors from which they are synthesized. Tryptophan, Tyrosine, Lysine, and Ornithine are amino acid precursors, which undergo enzymatic reactions giving rise to indole, tetrahydroisoquinoline, piperidine, and pyrrolizidine alkaloids, respectively). In addition to nitrogenous bases, serotonin derivatives like N-feruloyl serotonin and N-(p-coumaroyl) serotonin, and spermidines have also been isolated from safflower. Based on their inhibitory action against melanin production, it was suggested that they can be potential inhibitors of melanogenesis (Zhou et al. 2014). Protective effects of the serotonin derivatives against cardiovascular diseases were investigated in ischemic and reperfused heart of guinea-pig. The study showed that there was improvement in relevant biological markers after administration of the two alkaloids and played important antioxidant roles for such conditions (Al-Snafi 2015).

3.2 Genetic Improvement of Non-seed Based Products from Safflower

Although safflower is used widely in Chinese medicine, the major attention for the breeder has been the safflower oil due to its health-related and commercial benefits. Thus, in contrast to seed-based products, very few genetic studies have been done for non-seed based products. However, some work has been done on flavonoids through genomics-aided breeding.

Transcriptomic studies including full-length transcriptome sequence analysis for the major flavonoid – HYSA, has shown that *CtC4H2, CtCHS3, CtCHI3, CtF3H3*, and *CtF3H1* are the major genes that are involved in flavonoid production (Chen et al. 2018). These known putative flavonoid genes can be the basis for biotechnological improvement of safflower. In a study by Yang et al. (2007), cDNA-AFLP was used in combination with bulked segregant analysis to mine differentially expressed genes associated with high HSYA content. Transcript derived fragments (TDFs) which showed significant association with HSYA content were retained for construction of linkage maps. Genetic linkage analysis showed that TDF-2, TDF-3, and TDF-9 were tightly linked to the HSYA genomic region. Another key regulatory enzyme of the flavonoid pathway is Chalcone isomerase (CHI). Guo et al. (2019) successfully obtained transgenic *Arabidopsis* and Safflower lines overexpressing *Ct*CHI1 gene. The overexpressed *Ct*CHI1 genes caused the 3.9 fold upregulation of*Ct*PAL3 and *Ct*C4H1 genes and downregulation/inhibition of *Ct*4CL3, *Ct*F3H, and *Ct*DFR2. Also, the comparison of transgenic and control groups revealed that there were 788 different metabolites marked and mostly were upregulated. Wang (2015) achieved overexpression of *CtAK* (key regulator of Aspartate metabolism) in transgenic safflower using CaMV35S promoter. Exogenous applications of ethylene and 1-aminocyclopropane carboxylic acid oxidase (ACO) were shown to affect the accumulation of flavonoids in safflower (Tu et al. 2019). Two of the ACO genes were cloned from safflower (*CtACO1* and *CtACO2*) and transgenic plants were developed through *Agrobacterium*-mediated floral dip method. In overexpressed lines of *CtACO1*, metabolite analysis showed the accumulation of quercetin and its glycosylated derivatives like rutin, while the amount of quinochalcones, kaempferol derivatives, apigenin, and luteolin were reduced.

Genetic studies were also done for the domestication-related traits in safflower. Sixty-one Quantitative Trait Loci (QTL) in various linkage groups were identified including a large-effect QTL corresponding to the flower colour (Pearl et al. 2014). A 3:1 segregation pattern was observed for flower colour in the F2 mapping population consisting of 276 individuals. This indicated that the changes in carthamin (the quinochalcone pigment responsible for the red-colored florets) synthesis are influenced by a single locus. These findings differed from other studies (Pahlavani et al. 2004) that showed the impact of multiple genes on flower color and the presence of at least two interacting genes that distinguishes the orange and the yellow-colored florets. These contrasting results are obtained probably due to selection of high carthamine-producing parents than any other compounds controlling the flower coloration to develop the mapping population by Pearl et al. (2014).

4 Social, Economic, and Political Impact of Nutraceutical and Pharmaceuticals

Healthy diet and lifestyle are key factors that influence prevention of diseases and promote betterment of human health. In recent years, there is a strong and increasing demand for development of functional food markets that mainly focus on increased production of nutraceuticals using organic and natural ingredients that are high on nutrition (Daliri and Lee 2015).

Safflower offers good raw material for several economic sectors. It is widely cultivated because of its ability to survive under varying edaphoclimatic conditions (Menegaes and Nunes 2020). Safflower oil is rich in linoleic acid containing tocopherols that act as antioxidants. It is recommended to patients suffering from diabetes and cardiovascular diseases, as it helps in reducing the level of cholesterol in blood similar to olive oil (Menegaes and Nunes 2020).

Due to advances in agricultural biotechnology, it is now possible to produce food crops which have enhanced nutritional content and can contribute towards overall human health. Biofortification practices are generally employed in order to enhance the nutritional value of food crops either by conventional breeding or through the use of biotechnology. Through these practices vitamins and minerals have been added to crops in order to produce a diet that has well balanced nutrition and supplements to address the issues of a malnourished population successfully. Nutraceuticals include bioactive components such as polyunsaturated fatty acids (PUFA), antioxidants, phytochemicals that are found in plants. Genetic engineering is generally employed to improve natural and therapeutic values of natural foods (Prabavathy et al. 2022).

Nutraceutical sector includes three main divisions which include herbal/natural products, dietary supplements, and functional foods (Asif 2019). Globally, of these three, dietary supplements and herbal/natural products were the most rapidly growing divisions with 19.5 % and 11.6% of growth per year, respectively (Asif 2019). It has been observed that the use of genetically modified products, increased income of farmers by $92 million between 1996 and 2011 because the overall production cost got decreased, and there was reduction in crop loss because GM crops were insect and pest resistant (Karalis et al. 2020). The Nutraceutical market has grown exponentially from the last decade in developed and developing nations all over the world. Globally, USA and Japan have well established nutraceutical markets and their income from this sector is showing a consistent increase. Among developing nations, India, China, and Brazil are taking the lead in the field of nutraceuticals. India and China are now providing the raw sources for development of natural products. Nutraceutical foods were the major market section with worth of 39.9 billion USD in year 2007 and increased up to 56.7 billion USD by 2013 (Singh et al. 2019). Beverages included under nutraceuticals had worth of 38.4 billion USD in 2007 and 71.3 billion USD in 2013. The nutraceutical markets grew by around 250 billion USD in 2018, and it has been observed that the evolution of nutraceuticals within these years has helped to improve the average lifespan of adult people specially belonging to age of more than 60 years (Bhowmik et al. 2013). In 2015, the estimated value of the nutraceuticals market in India was 4 billion

USD however, it is expected to increase to 10 billion USD by 2022. Ministry of AYUSH and companies like Dabur, Patanjali, etc., have played a significant role in expanding the nutraceutical market in recent years (Singh et al. 2019).

Under the Food Safety and Standards Act, 2006, a regulatory body has been established which was named as Food Safety and Standards Authority of India (FSSAI). Through combining the various acts and orders under different governmental bodies, it manages the food related issue in India. FSSAI functions to ensure that safe and wholesome food is available for human consumption and regulates import, distribution, manufacture, storage, and sale of nutraceuticals and dietary supplements (Verma and Popli 2018). In USA, nutraceuticals and functional foods are tightly monitored by the Federal Food, Drug, and Cosmetic Act by the Food and Drugs Authority (FDA), USA. Safety and labelling of dietary supplements is also under the control of FDA to ensure that they are able to fulfil the needs of FDA and Dietary Supplement Health and Education (DSHE) regulations before being marketed (Daliri and Lee 2015). Before releasing new supplements into the markets, manufacturers are required to inform FDA and provide information indicating that their product qualifies as a new dietary supplement and is safe to use as specified in labelling. The Nutraceutical industry has to face a lot of challenges including adherence to country-specific regulations, less innovation by food industries (Daliri and Lee 2015). When a company obtains the petition for health claim and receives the acceptance, other competitors can also use their claim since majority of companies use products containing "easily available" ingredients which can be copied, unless and until the nutraceutical products contain exclusive ingredients that may be patented. The Nutraceutical sector can achieve more milestones when food companies become more innovative. Currently, they are functionalizing conventional products by adding more vitamins, herbal extracts, and minerals rather than inventing new products.

5 Future Prospects

Being rich in nutra and pharmaceutical compounds, safflower has a scientific and economic rationale for its use in the nutraceutical industry. Its rich nutritional value can also provide food security and improve the economy of developing countries. Many countries have adopted safflower as a mainstream crop and are now among the top producers in terms of production and acreage. Due to its richness of oleic acid and linoleic acid, Safflower has been used widely for its oil. Being a semi-arid crop, it can grow in the area where irrigation is limited. Safflower is also known for various important medicinal and nutritional benefits, which makes it an ideal candidate for use as a nutraceutical as well as pharmaceutical. However, the knowledge accumulation has been slow as safflower is a marginal crop, but it promises much more due to the benefits it brings along. With the availability of genome sequence data and sufficient genetic variability, the crop requires concentrated efforts to boost its status as a mainstream crop in nutraceutical and pharmaceutical industry.

References

Abdurakhmonov IY, Abdukarimov A (2008) Application of association mapping to understanding the genetic diversity of plant germplasm resources. Int J Plant Genom 2008:1–18

Abidi SL (2001) Chromatographic analysis of plant sterols in foods and vegetable oils. J Chromatogr A 935:173–201

Almeida R, Navarro D, Barbosa-Filho J (2001) Plants with central analgesic activity. Phytomedicine 8:310–322

Al-Snafi AE (2015) The chemical constituents and pharmacological importance of Carthamus tinctorius-an overview. J Pharm Biol 5:143–166

Ambreen H, Kumar S, Variath MT, Joshi G, Bali S, Agarwal M, Kumar A, Jagannath A, Goel S (2015) Development of genomic microsatellite markers in Carthamus tinctorius L.(safflower) using next generation sequencing and assessment of their cross-species transferability and utility for diversity analysis. PLoS One 10:e0135443

Ambreen H, Kumar S, Kumar A, Agarwal M, Jagannath A, Goel S (2018) Association mapping for important agronomic traits in safflower (Carthamus tinctorius L.) core collection using microsatellite markers. Front Plant Sci 9:402

Amini F, Saeidi G, Arzani A (2008) Study of genetic diversity in safflower genotypes using agromorphological traits and RAPD markers. Euphytica 163:21–30

Arslan B (2007) The path analysis of yield and its components in safflower (Carthamus tinctorius L.). J Biol Sci 7:668–672

Asgary S, Rahimi P, Mahzouni P, Madani H (2012) Antidiabetic effect of hydroalcoholic extract of Carthamus tinctorius L. in alloxan-induced diabetic rats. J Res Med Sci 17:386

Ashri A, Knowles P (1960) Cytogenetics of safflower (Carthamus L.) species and their hybrids 1. Agron J 52:11–17

Ashri A, Zimmer D, Urie A, Cahaner A, Marani A (1974) Evaluation of the world collection of safflower, Carthamus tinctorius L. IV. Yield and yield components and their relationships 1. Crop Sci 14:799–802

Asif M, Mohd I (2019) Prospects of medicinal plants derived nutraceuticals: a re-emerging new era of medicine and health aid. Progress Chem Biochem Res 2:150–169

Bae SJ, Shim SM, Park YJ, Lee JY, Chang EJ, Choi SW (2002) Cytotoxicity of phenolic compounds isolated from seeds of safflower (Carthamus tinctorius L.) on cancer cell lines. Food Sci Biotechnol 11:140–146

Bergman J, Carlson G, Kushnak G, Riveland N, Stallknecht G (1985) Registration of 'Oker'safflower. Crop Sci 25:1127–1128

Bowers JE, Pearl SA, Burke JM (2016) Genetic mapping of millions of SNPs in safflower (Carthamus tinctorius L.) via whole-genome resequencing. G3: genes. Genom Genet 6:2203–2211

Carlsson AS, Zhu L-H, Andersson M, Hofvander P (2014) Platform crops amenable to genetic engineering–a requirement for successful production of bio-industrial oils through genetic engineering. Biocatal Agric Biotechnol 3:58–64

Chakradhari S, Perkons I, Mišina I, Sipeniece E, Radziejewska-Kubzdela E, Grygier A, Rudzińska M, Patel KS, Radzimirska-Graczyk M, Górnaś P (2020) Profiling of the bioactive components of safflower seeds and seed oil: cultivated (Carthamus tinctorius L.) vs. wild (Carthamus oxyacantha M. Bieb.). Eur Food Res Technol 246:449–459

Chen Y-S, Lee S-M, Lin C-C, Liu C-Y, Wu M-C, Shi W-L (2013) Kinetic study on the tyrosinase and melanin formation inhibitory activities of carthamus yellow isolated from Carthamus tinctorius L. J Biosci Bioeng 115:242–245

Cho S-H (2001) Effects of defatted safflower and perilla seed powders on lipid metabolism in ovariectomized female rats fed high cholesterol diets. J Korean Soc Food Sci Nutr 30:112–118

Collard BC, Jahufer M, Brouwer J, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142:169–196

Dajue L, Mündel H-H (1996) Promoting the conservation and use of underutilized and neglected crops. Safflower 7:9–38

Dureja H, Kaushik D, Kumar V (2003) Developments in nutraceuticals. Indian J Pharmacol 35: 363–372

Ebert W, Knowles P (1966) Inheritance of pericarp types, sterility, and dwarfhess in several safflower crosses 1. Crop Sci 6:579–582

Fernandez-Martinez J, Del Rio M, De Haro A (1993) Survey of safflower (Carthamus tinctorius L.) germplasm for variants in fatty acid composition and other seed characters. Euphytica 69:115–122

García-Moreno MJ, Fernández-Martínez JM, Velasco L, Pérez-Vich B (2011) Molecular tagging and candidate gene analysis of the high gamma-tocopherol trait in safflower (Carthamus tinctorius L.). Mol Breed 28:367–379

Gecgel U, Demirci M, Esendal E, Tasan M (2007) Fatty acid composition of the oil from developing seeds of different varieties of safflower (Carthamus tinctorius L.). J Am Oil Chem Soc 84:47–54

Ghareghani M, Zibara K, Azari H, Hejr H, Sadri F, Jannesar R, Ghalamfarsa G, Delaviz H, Nouri E, Ghanbari A (2017) Safflower seed oil, containing oleic acid and palmitic acid, enhances the stemness of cultured embryonic neural stem cells through Notch1 and induces neuronal differentiation. Front Neurosci 11:446

Golkar P, Arzani A, Rezaei A (2011) Determining relationships among seed yield, yield components and morpho-phenological traits using multivariate analyses in safflower (Carthamus tinctorious L.). Ann Biol Res 2:162–169

Golkar P, Arzani A, Rezaei A (2012) Evaluation of genetic control of seed quality traits in safflower (Carthamus tinctorious L.). Iranian J Field Crop Sci 42:833–844

Guo D, Xue Y, Li D, He B, Jia X, Dong X, Guo M (2017) Overexpression of CtCHS1 increases accumulation of quinochalcone in safflower. Front Plant Sci 8:1409

Hamdan Y, Pérez-Vich B, Fernández-Martínez J, Velasco L (2008) Inheritance of very high linoleic acid content and its relationship with nuclear male sterility in safflower. Plant Breed 127:507–509

Hamdan YA, Pérez-Vich B, Velasco L, Fernández-Martínez JM (2009a) Inheritance of high oleic acid content in safflower. Euphytica 168:61–69

Hamdan YA, Pérez-Vich B, Fernández-Martínez JM, Velasco L (2009b) Novel safflower germplasm with increased saturated fatty acid content. Crop Sci 49:127–132

Hamdan YA, García-Moreno MJ, Fernández-Martínez JM, Velasco L, Pérez-Vich B (2012) Mapping of major and modifying genes for high oleic acid content in safflower. Mol Breed 30:1279–1293

Hassani SMR, Talebi R, Pourdad SS, Naji AM, Fayaz F (2020) Morphological description, genetic diversity and population structure of safflower (Carthamus tinctorius L.) mini core collection using SRAP and SSR markers. Biotechnol Biotechnol Equipment 34:1043–1055

Hattori M, Xin-Li H, Qing-Ming C, Kawata Y, Tezuka Y, Kikuchi T, Namba T (1992) 6-Hydroxykaempferol and its glycosides from Carthamus tinctorius petals. Phytochemistry 31:4001–4004

Huang Z, Yu M, Qu S, Shan M, Song C (1984) Studies on the immuno-activity of the polysaccharide from safflower (Carthamus tinectorius). Chin Trad Herb Drugs 15:213–216

Huang J, Yang J, Guan L, Yi S, Du L, Tian H, Guo Y, Zhai F, Lu Z, Li H (2017) Expression of bioactive recombinant human fibroblast growth factor 10 in Carthamus tinctorius L. seeds. Protein Expr Purif 138:7–12

Iwata T, Hoshi S, Takehisa F, Tsutsumi K, Furukawa Y, Kimura S (1992) The effect of dietary safflower phospholipid and soybean phospholipid on plasma and liver lipids in rats fed a hypercholesterolemic diet. J Nutr Sci Vitaminol 38:471–479

Jiang Z, Wang J-D, Chen N-S, Huang J-L (2008) 2-Benzylisoindoline-1, 3-dione: a monoclinic polymorph. Acta Crystallogr Sect E: Struct Rep Online 64:o324–o324

Johnson R, Bergman J, Flynn C (1999) Oil and meal characteristics of core and non-core safflower accessions from the USDA collection. Genet Resour Crop Evol 46:611–618

Kadirvel P, Veerraju C, Senthilvel S, Yadav P, Kiran BU, Shaik M, Shaw R, Selvaraj VM, Reddy YR, Mohanrao MD (2020) Marker-assisted selection for fast-track breeding of high oleic lines in safflower (Carthamus tinctorious L.). Ind Crop Prod 158:112983

Kanehira T, Saito K (1990) Decomposition of carthamin by peroxidases from Carthamus tinctorius. Biochem Physiol Pflanz 186:179–187

Karalis DT, Karalis T, Karalis S, Kleisiari AS (2020) Genetically modified products, perspectives and challenges. Cureus 12

Karimi S (2015) Study of physiological traits and microsatellite markers associated with salt tolerance in safflower (*Carthamus tinctorius* L.) Dissertation, Isfahan University of Technology

Kazuma K, Takahashi T, Sato K, Takeuchi H, Matsumoto T, Okuno T (2000) Quinochalcones and flavonoids from fresh florets in different cultivars of Carthamus tinctorius L. Biosci Biotechnol Biochem 64:1588–1599

Kim M, Le Scao-Bogaert F, Paris M (1992) Flavonoids from Carthamus tinctorius flowers. Planta Med 58:285–286

Knowles PF (1989) In: Downey RK, Robbelen G, Ashri A (eds) Safflower: Oil crops of thw World. McGraw-Hill, New York, pp 363–374

Koyama N, Kuribayashi K, Seki T, Kobayashi K, Furuhata Y, Suzuki K, Arisaka H, Nakano T, Amino Y, Ishii K (2006) Serotonin derivatives, major safflower (Carthamus tinctorius L.) seed antioxidants, inhibit low-density lipoprotein (LDL) oxidation and atherosclerosis in apolipoprotein E-deficient mice. J Agric Food Chem 54:4970–4976

Kumar S, Ambreen H, Variath MT, Rao AR, Agarwal M, Kumar A, Goel S, Jagannath A (2016) Utilization of molecular, phenotypic, and geographical diversity to develop compact composite core collection in the oilseed crop, safflower (Carthamus tinctorius L.) through maximization strategy. Front Plant Sci 7:1554

Lee JY, Chang EJ, Kim HJ, Park JH, Choi SW (2002) Antioxidative flavonoids from leaves of Carthamus tinctorius. Arch Pharm Res 25:313–319

Li D, Wang Q, Xu X, Yu J, Chen Z, Wei B, Wu W (2021) Temporal transcriptome profiling of developing seeds reveals candidate genes involved in oil accumulation in safflower (Carthamus tinctorius L.). BMC Plant Biol 21:1–17

Liu Q, Cao S, Zhou X-R, Wood C, Green A, Singh S (2013) Nonsense-mediated mRNA degradation of CtFAD2-1 and development of a perfect molecular marker for olol mutation in high oleic safflower (Carthamus tinctorius L.). Theor Appl Genet 126:2219–2231

Mani V, Lee S-K, Yeo Y, Hahn B-S (2020) A metabolic perspective and opportunities in pharmacologically important safflower. Meta 10:253

Mayerhofer R, Archibald C, Bowles V, Good AG (2010) Development of molecular markers and linkage maps for the Carthamus species C. tinctorius and C. oxyacanthus. Genome 53:266–276

Menegaes JF, Nunes UR (2020) Safflower: importance, use and economical exploitation. Sci Agraria Paranaensis:1–11

Mensink RP, Katan MB (1989) Effect of a diet enriched with monounsaturated or polyunsaturated fatty acids on levels of low-density and high-density lipoprotein cholesterol in healthy women and men. N Engl J Med 321:436–441

Meselhy MR, Kadota S, Momose Y, Hatakeyama N, Kusai A, Hattori M, Namba T (1993) Two new quinochalcone yellow pigments from Carthamus tinctorius and Ca2+ antagonistic activity of tinctormine. Chem Pharm Bull 41:1796–1802

Mirzahashemi M, Mohammadi-Nejad G, Golkar P (2015) A QTL linkage map of safflower for yield under drought stress at reproductive stage. Iranian J Genet Plant Breeding 4:20–27

Mizukami H, Inagaki C, Okabe Y, Okuyama H (2000) cDNA cloning and characterization of a novel gene differentially expressed in developing seeds of high-oleate safflower (Carthamus tinctorius L.). Plant Biotechnol 17:315–319

Nykiforuk CL, Shen Y, Murray EW, Boothe JG, Busseuil D, Rheaume E, Tardif JC, Reid A, Moloney MM (2011) Expression and recovery of biologically active recombinant Apolipoprotein AIMilano from transgenic safflower (Carthamus tinctorius) seeds. Plant Biotechnol J 9:250–263

Nykiforuk CL, Shewmaker C, Harry I, Yurchenko OP, Zhang M, Reed C, Oinam GS, Zaplachinski S, Fidantsef A, Boothe JG (2012) High level accumulation of gamma linolenic acid (C18: 3Δ6. 9, 12 cis) in transgenic safflower (Carthamus tinctorius) seeds. Transgenic Res 21:367–381

Pearl SA, Bowers JE, Reyes-Chin-Wo S, Michelmore RW, Burke JM (2014) Genetic analysis of safflower domestication. BMC Plant Biol 14:1–15

Peng X-R, Wang X, Dong J-R, Qin X-J, Li Z-R, Yang H, Zhou L, Qiu M-H (2017) Rare hybrid dimers with anti-acetylcholinesterase activities from a safflower (Carthamus tinctorius L.) seed oil cake. J Agric Food Chem 65:9453–9459

Porokhovinova E, Matveeva T, Khafizova G, Bemova V, Doubovskaya A, Kishlyan N, Podolnaya L, Gavrilova V (2022) Fatty acid composition of oil crops: genetics and genetic engineering. Genet Res Crop Evolut 69:1–17

Prabavathy D, Sudha S, Inbathamizh L (2022) Genetically modified products and non-GMO products in nutraceuticals. In: Handbook of nutraceuticals and natural products: biological, medicinal, and nutritional properties and applications, vol 2. John Wiley & Sons, Inc., Hoboken, pp 315–336

Pritchard F, Eagles H, Norton R, Salisbury P, Nicolas M (2000) Environmental effects on seed composition of Victorian canola. Aust J Exp Agric 40:679–685

Rahimi P, Asgary S, Kabiri N (2014) Hepatoprotective and hypolipidemic effects of Carthamus tinctorius oil in alloxan-induced type 1 diabetic rats. J HerbMed Pharmacol 3:107–111

Sakamura S, Terayama Y, Kawakatsu S, Ichihara A, Saito H (1978) Conjugated serotonins related to cathartic activity in safflower seeds (Carthamus tinctorius L.). Agric Biol Chem 42:1805–1806

Shirwaikar A, Khan S, Kamariya YH, Patel BD, Gajera FP (2010) Medicinal plants for the management of post menopausal osteoporosis: a review. Open Bone J 2

Sun Y, Guo T, Sui Y, Li F (2003) Quantitative determination of rutin, quercetin, and adenosine in Flos Carthami by capillary electrophoresis. J Sep Sci 26:1203–1206

Takahashi T, Miyazawa M (2012) Potent α-glucosidase inhibitors from safflower (Carthamus tinctorius L.) seed. Phytother Res 26:722–726

Takahashi Y, Miyasaka N, Tasaka S, Miura I, Urano S, Ikura M, Hikichi K, Matsumoto T, Wada M (1982) Constitution of two coloring matters in the flower petals of Carthamus tinctorius L. Tetrahedron Lett 23:5163–5166

Takahashi Y, Saito K, Yanagiya M, Ikura M, Hikichi K, Matsumoto T, Wada M (1984) Chemical constitution of safflor yellow B, a quinochalcone c-glycoside from the flower petals of carthamus tinctorius L. Tetrahedron Lett 25:2471–2474

Topkafa M (2016) Evaluation of chemical properties of cold pressed onion, okra, rosehip, safflower and carrot seed oils: triglyceride, fatty acid and tocol compositions. Anal Methods 8:4220–4225

Urie A (1986) Inheritance of partial hull in safflower 1. Crop Sci 26:493–498

Velasco L, Fernández-Martínez J (2001) Breeding for oil quality in safflower. Proceedings of the 5th International Safflower Conference, Williston, North Dakota and Sidney, Montana, USA, 23–27 July, 2001 Safflower: a multipurpose species with unexploited potential and world adaptability. Department of Plant Pathology, North Dakota State University, pp. 133-137

Velasco L, Pérez-Vich B, Fernández-Martínez J (2005) Identification and genetic characterization of a safflower mutant with a modified tocopherol profile. Plant Breed 124:459–463

Verma B, Popli H (2018) Regulations of nutraceuticals in India & US. Pharm Innovat J 7:811–816

Vieira MG, Batista LR, Muniz AS, Lichston JE, Oliveira AR, César-Oliveira MA, Antoniosi Filho NR (2021) Synthesis of antioxidant additive from safflower seed oil. J Braz Chem Soc 32:40–46

Vosoughkia M, Hossainchi GL, Ghavami M, Gharachorloo M, Delkhosh B (2011) Evaluation of oil content and fatty acid composition in seeds of different genotypes of safflower (Carthamus tinctorius L.)

Wang Y-F (2015) Molecular cloning, sequence analysis, and construction of expression vector of monofunction feedback inhibition sensitive CtAK1 gene in Carthamus tinctorius. Chin Tradit Herb Drugs:3740–3745

Watanabe T, Terabe S (2000) Analysis of natural food pigments by capillary electrophoresis. J Chromatogr A 880:311–322. a. action. Neurosci Lett 386:58–62

Wood CC, Okada S, Taylor MC, Menon A, Mathew A, Cullerne D, Stephen SJ, Allen RS, Zhou XR, Liu Q (2018) Seed-specific RNAi in safflower generates a superhigh oleic oil with extended oxidative stability. Plant Biotechnol J 16:1788–1796

Wu Z, Liu H, Zhan W, Yu Z, Qin E, Liu S, Yang T, Xiang N, Kudrna D, Chen Y (2021) The chromosome-scale reference genome of safflower (Carthamus tinctorius) provides insights into linoleic acid and flavonoid biosynthesis. Plant Biotechnol J 19:1725–1742

Yang Y-X, Wu W, Zheng Y-L, Chen L, Liu R-J, Huang C-Y (2007) Genetic diversity and relationships among safflower (Carthamus tinctorius L.) analyzed by inter-simple sequence repeats (ISSRs). Genet Resour Crop Evol 54:1043–1051

Yue S, Tang Y, Li S, Duan J-A (2013) Chemical and biological properties of quinochalcone C-glycosides from the florets of Carthamus tinctorius. Molecules 18:15220–15254

Zang B, Jin M, Si N, Zhang Y, Wu W, Piao Y (2002) Antagonistic effect of hydroxysafflor yellow A on the platelet activating factor receptor. Yao Act Pharm Sin 37:696–699

Zhang Z, Johnson R (1999) Safflower germplasm collection directory. IPGRI Office for East Asia International Plant Genetic Resources Institute, Rome

Zhang HL, Nagatsu A, Watanabe T, Okuyama H (1997) Antioxidative compounds isolated from safflower (Carthamus tinctorius L.) oil cake. Chem Pharm Bull 45:1910–1914

Zhang Z, Li Q, Liu F, Sun Y, Zhang J (2010) Prevention of diet-induced obesity by safflower oil: insights at the levels of PPARα, Orexin, and Ghrelin gene expression of adipocytes in mice. Acta Biochim Biophys Sin 42:202–208

Zhang L-L, Tian K, Tang Z-H, Chen X-J, Bian Z-X, Wang Y-T, Lu J-J (2016) Phytochemistry and pharmacology of Carthamus tinctorius L. Am J Chin Med 44:197–226

Zhou Y-Z, Ma H-Y, Chen H, Qiao L, Yao Y, Cao J-Q, Pei Y-H (2006) New acetylenic glucosides from Carthamus tinctorius. Chem Pharm Bull 54:1455–1456

Zhou Y-Z, Chen H, Qiao L, Xu N, Cao J-Q, Pei Y-H (2008) Two new compounds from Carthamus tinctorius. J Asian Nat Prod Res 10:429–433

Zhou X, Tang L, Xu Y, Zhou G, Wang Z (2014) Towards a better understanding of medicinal uses of Carthamus tinctorius L. in traditional Chinese medicine: a phytochemical and pharmacological review. J Ethnopharmacol 151:27–43

Oil Palm: Genome Designing for Improved Nutritional Quality

Maizura Ithnin, Abrizah Othman, Noor Idayu Mhd Tahir, Kalyana Babu Banisetti, Mohd Amin Abd Halim, and M. K. Rajesh

Contents

Abstract

Edible palm oil products are free from trans fats, making them healthier and safe for application in food industry. Palm oil contains valuable vitamins and phytonutrients that exhibit cardioprotective mechanisms, immune system enhancement, neurodegeneration protection, and antioxidative as well as anti-carcinogenic properties. Red palm oil (RPO), a component of crude palm oil, is beneficial as a vitamin A supplement for treating vitamin A deficiency and its related diseases. The nutritional value of palm oil can be further enhanced via conventional breeding, using, for instance, palms exhibiting high carotene and

M. Ithnin (✉) · A. Othman · N. I. M. Tahir · M. A. Abd Halim
Malaysian Palm Oil Board, Persiaran Institusi, Bandar Baru Bangi, Kajang, Malaysia
e-mail: maizura@mpob.gov.my

K. B. Banisetti
ICAR-Indian Institute of Oil Palm Research, Eluru, India

M. K. Rajesh
Division of Crop Improvement, ICAR-Central Plantation Crops Research Institute, Kasaragod, India

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_22

vitamin E contents that are available in the oil palm germplasm collection. Genes for key enzymes involved in carotene synthesis in palm oil have been isolated and characterized. These, together with the oil palm genome builds and its associated databases, provide resources for developing marker-assisted selection (MAS) programs and support genetic engineering technologies toward improving the nutritional values of oil palm and palm oil.

Keywords

Palm oil · Phytonutrients · Red palm oil · Vitamin E · Nutritional genomics

1 Agricultural Importance of the Oil Palm Crop

Oil palm is a perennial tree crop that produces two types of oils, namely, palm oil which is extracted from the fleshy mesocarp and palm kernel oil which is obtained from the oil palm kernel. The former has vast applications in the food industry, while the latter is mostly used in the oleochemical industry. Oil palm is a monocotyledon classified under the genus *Elaeis*. This genus consists of two species, *Elaeis guineensis* and *Elaeis oleifera*. *Elaeis guineensis* originates from Central and West Africa and is presently the main oil palm planting material planted commercially. *E. oleifera* has a South and Central America origin and possesses interesting features such as high carotene content (Mohd Din et al. 2002), low height increment rate, high oil unsaturation (Hardon 1969), and tolerance to diseases (Turner 1981), which are pertinent to oil palm improvement.

The oil palm has the highest productivity among oil-producing plants, producing approximately 3.7 tons of oil hectare^{-1} year^{-1}. With 11 and 10 times the yield capacity of soybean and rapeseed, respectively (Khosla and Sundram 2010), oil palm produces 75 million tons of oil from only ~24 million hectares of planted area. This output is obtained from less than 5% of the total area planted with oil crops. For many years, palm oil has contributed approximately 30% of the total oils and fats produced worldwide (Oil World 2022). Palm oil recorded a 53.4% market share of the major global oil and fat exports in 2021 (Oil World 2022). The main producers of palm oil are Indonesia (58.8%), Malaysia (23.8%), and South and Central America (6.9%). Palm oil is a rich source of vitamins and phytonutrients. The abundance of worldwide supplies and vast applications in the food industry make palm oil a prospective source of vitamins and phytonutrients that could meet the micronutrient requirements across the globe.

2 Chemical Composition of Palm Oil and Oil Palm

Palm oil and palm kernel oil differ in terms of their fatty acid compositions, particularly in the proportions of saturated fatty acids (SFA) and unsaturated fatty acids (UFA). Palm kernel oil has properties similar to coconut oil, with 82% SFA and

18% UFA. The 82% SFA is mainly composed of lauric (~50%), myristic (~16%), and palmitic acids (~8%), while the 18% UFA includes oleic acid (15%) and linoleic acid (~3%) (Ibrahim 2013). On the other hand, palm oil contains balanced amounts of SFA and UFA. The main SFAs in palm oil are palmitic acid (45%) and stearic acid (5%). Oleic acid and linoleic acid are the UFA present in palm oil at 40% and 10%, respectively. Besides fatty acids, palm oil also contains a wide range of health-benefiting phytonutrients such as tocopherols, tocotrienols, carotenoids, polyphenols, phytosterols, squalene, phospholipids, and coenzyme Q_{10} or ubiquinone-10 (Choo and Nesaretnam 2014). The detailed fatty acid composition of palm oil and palm kernel oil is listed in Table 1.

There are also active research and development (R&D) programs to improve and innovate the milling and processing of palm oil to preserve the natural phytonutrients. The phytonutrients in red or cold-pressed palm oil, for instance, are extracted for food, cosmetic, and pharmaceutical applications (Hassan et al. 2021; Abd Rashid et al. 2021). Red palm oil (RPO) contains vitamin E (α-, β-, and γ-tocotrienols), carotenoids (xanthophylls, α- and β-carotenes), phytosterols, ubiquinone, and squalene and is a competitive and better alternative to other consumer oils due to its nutritional content (Loganathan et al. 2017). Oil palm-based products and by-products from the fruit bunches, kernel shells, trunk, fronds, and pressed fruits also contain phenolic compounds, terpenes, lignin, lignans, vitamins, sugar, and minerals (Ofori-Boateng and Lee 2013). The aqueous waste from the palm fruit milling and processing has been extensively studied and valued for its high content of phenolic compounds (Syarifah-Noratiqah et al. 2019).

Table 1 Palm oil and palm kernel oil fatty acid composition

Fatty acid	Lipid number, carbon/double bond (C:D)	Saturation	Palm oil (%)	Palm kernel oil (%)
Palmitic acid	16:0	Saturated	39.2–45.8	7.5–9.3
Oleic acid	18:1	Unsaturated	37.4–44.1	13.7–17.0
Linoleic acid	18:2	Unsaturated	8.7–12.5	2.1–2.9
Stearic acid	18:0	Saturated	3.7–5.4	1.8–2.4
Myristic acid	14:0	Saturated	0.9–1.5	15.4–17.2
Arachidic acid	20:0	Saturated	0.0–0.5	0.0–0.1
Linolenic acid	18:3	Unsaturated	0.0–0.6	–
Lauric acid	12:0	Saturated	0.0–0.5	45.4–49.8
Palmitoleic acid	16:1	Unsaturated	0.0–0.4	
Capric acid	10:0	Saturated	–	2.9–3.7
Caprylic acid	8:0	Saturated	–	3.2–4.7
Caproic acid	6:0	Saturated	–	0.2–0.4

Source: Bustamam et al. (2019), Japir et al. (2017), Mancini et al. (2015), Choo and Nesaretnam (2014), Ibrahim (2013)

3 Oil Palm Phytochemicals

The natural chemical components of plants are known as phytochemicals. The term "phyto" comes from a Greek word that means plant (Liu 2004). Phytochemicals are of fundamental biological importance to plants; they provide protection against pathogens, herbivores, and solar radiation and function as signaling molecules and regulators (Csepregi and Hideg 2017; Winkel 2004). Phytochemicals are known to be bioactive and antioxidative against free radicals, cancerous cells, bacteria, and viruses and have various therapeutic functions comprising anti-inflammatory, antimutagenic, and antitumor effects on human health (Ciucure and Geană 2019; Naithani et al. 2008). They are also the source of fine products for dyes, drugs, fragrances, and flavors which are important to the commercial industry (Shilpa et al. 2010).

Apart from the prized palm oil, the palm fruit contains phospholipids which are initially part of the cell membrane. Phospholipids are amphiphilic and are a source of lecithin and an important releasing agent in food products (Choo et al. 2004). Crude palm oil contains very low amounts of phospholipids due to the milling process, and, therefore, phospholipids are better extracted from the palm fruit fiber and calyx (Gold et al. 2016; Choo et al. 2004). The primary phospholipids in palm oil are phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylinositol (PI), and phosphatidylglycerol (PG) with a small amount of phosphatidic acid (PA), diphosphatidylglycerol (DPG), lysophosphatidylethanolamine (LPE), lysophosphatidylcholine (LPC), and phosphatidylserine (PS) (Goh et al. 1982). The fibrous by-products retained after oil extraction from the oil palm fruit, also known as oil palm fruit fiber, predominantly contain PC, PE, PG, and PA (Choo et al. 2004). Galactolipids of monogalactosyldiacylglycerol (MGDG) and digalactosyldiacylglycerol (DGDG), which are also part of palm fruit cell membranes, were profiled from palm oil (Cheong et al. 2014). These glycolipids are reported to show health-promoting effects (Christensen 2009).

Terpenes are natural unsaturated hydrocarbons that form the largest class of plant secondary metabolites. They are made up of five carbon isoprene units assembled into multiple pairs of isoprene pairs and various additions of side chains and functional groups (Perveen 2018). Terpenes can take linear or cyclic forms and are categorized according to their isoprene units and structural variations. Terpenoids are terpenes with additional, removal, or substitution of functional groups, for example, oxygen, methylene groups, or hydrogen atoms (Masyita et al. 2022). Terpenes include squalene, sterol, saponin, and carotenoid, while the terpene isoprene unit is the side chain of tocotrienols (Ahsan et al. 2015). Squalene is found in palm oil and palm oil milling waste. It is a biochemical intermediate in the synthesis of phytosterols in plants and hormones in humans. Squalene acts as an antioxidant and is an important ingredient especially in the cosmetic industry (Gonzalez-Diaz and García-Núñez 2021). Crude palm oil and its by-products contain phytosterols of β-sitosterol, campesterol, and stigmasterol (Jalani et al. 2021; Choo and Nesaretnam 2014). These plant sterols modulate membrane-bound enzyme activities and demonstrate cholesterol-lowering activity (De Smet et al. 2012).

Derived from tetraterpenes, carotenoids can be classified into xanthophylls and carotenes based on their oxygen content (Cazzaniga et al. 2016). Xanthophylls of

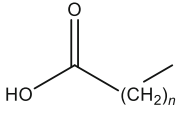
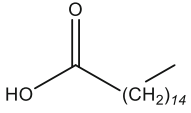
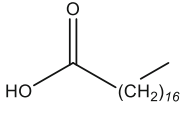
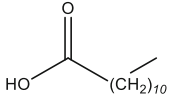
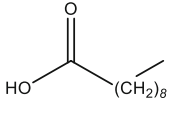
lutein and carotene epoxides and at least 11 other types of carotenes were profiled in palm oil, namely, phytoene, phytofluene, α-carotene, β-carotene, ξ-carotene, γ-carotene, δ-carotene, lycopene, α-zeacarotene, β-zeacarotene, and neurosporene (Ping and Gwendoline 2006; Ng and Choo 2016). Tocopherols and tocotrienols are collectively identified as vitamin E and are important palm oil constituents. Crude palm oil, palm olein, and red palm olein were all reported to retain α-tocopherol (α-T), α-tocotrienol (α-T3), γ-tocotrienol (γ-T3), and δ-tocotrienol (δ-T3).

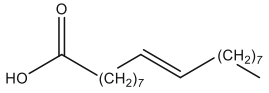
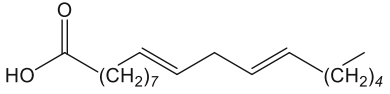
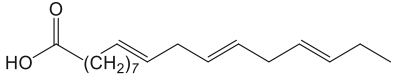
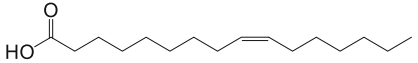
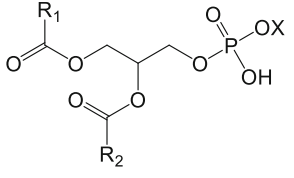
Lignans are water-soluble polyphenols formed from monolignols of coumaryl, coniferyl alcohol, or sinapyl alcohols. On the other hand, lignin is synthesized from similar monolignols but is linked differently by dirigent proteins resulting in hydrophobic natural polymers in the plant cell wall (Davin and Lewis 2000). Isolariciresinol, lariciresinol, and pinoresinol are examples of lignans detected in palm oil (Tardugno et al. 2022). These lignans exert beneficial effects on human health (Rodríguez-García et al. 2019). Other secondary metabolites such as phenolic compounds are detected in the by-products of the milling process due to their low lipophilic nature. The term "phenolic" includes metabolites with hydroxyl-substituted benzene ring synthesized by the phenylpropanoid biosynthetic pathway. Flavonoids, the largest group of phenolics, are composed of a diphenyl-propane (C6-C3-C6) backbone in which two aromatic rings are connected via a three-carbon chain (Alseekh et al. 2020). Flavonoids occur naturally as aglycone or as conjugates with sugars, organic acids, and other molecules and are present in the oil palm fruit and leaf (Hazir et al. 2012; Tahir et al. 2013).

Organic acids in plants typically form alkaline salt, ester, and glycoside that are very important in human nutrition as biochemical pathway intermediaries (Gundogdu et al. 2014), while amino acids are the building blocks for tissue proteins and are substrates for the synthesis of numerous substances of physiological importance in the human body (Wu 2013). Table 2 explains each class of the nutritive phytochemicals characterized in palm oil and oil palm tissues and their biosynthesis pathway references from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000).

Other oil palm plant parts such as the leaves, fruits, and cabbages (palm heart) are mostly edible by humans and ruminants (Reddy et al. 2019; Ebrahimi et al. 2015). Alkaloids are bitter to the taste and can be found in plants in small quantities (Muñoz et al. 2020). A qualitative investigation revealed that alkaloids are the least phytochemical component found in oil palm leaves (Yin et al. 2013) and were not identified exhaustively except for β-phenylethylamine derivatives of tyramine and catecholamines in environment and disease studies (Tahir et al. 2022; Rodrigues-Neto et al. 2018). Catechins are a well-studied group of flavonoids found at high levels in tea and are also recorded in oil palm leaf and root (Tahir et al. 2013, 2022). Procyanidin B, the oligomer of catechin, is also found in oil palm root (Nurazah et al. 2013). Ferulic, sinapic, coumaric, and caffeic acids are among the hydroxycinnamic phenolic acids found in oil palm fruits (Sambanthamurthi et al. 2011). Dietary fibers and sugars derived from cellulosic and hemicellulosic components from oil palm biomass are also valuable and gaining attention from food and fuel industries (Mazlan et al. 2021; Palamae et al. 2017).

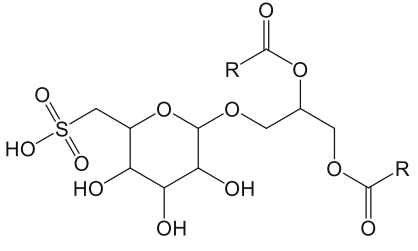
Table 2 Phytochemicals in oil palm products and by-products

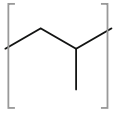
Class	Structure	Biosynthesis/pathway	References for oil palm
Palm oil			
Lipids			
Fatty acids		Fatty acid biosynthesis (KEGG pathway map00061)	Japir et al. (2017); Tahir et al. (2021)
Palmitic acid			
Stearic acid			
Lauric acid			
Capric acid			

Oleic acid			
Linoleic acid			
Linolenic acid			
Palmitoleic acid			
Phospholipids			
Glycerophospholipid (phosphoglyceride)		Glycerophospholipid metabolism (KEGG pathway map00564)	Gold et al. (2016); Choo et al. (2004)
	R = fatty acids		

(continued)

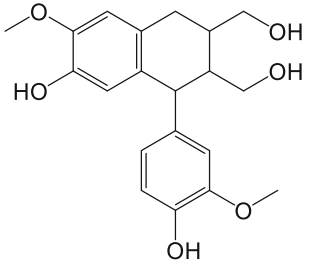
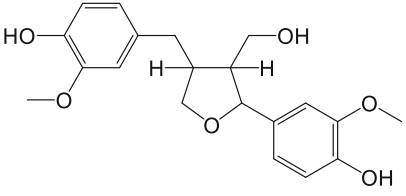
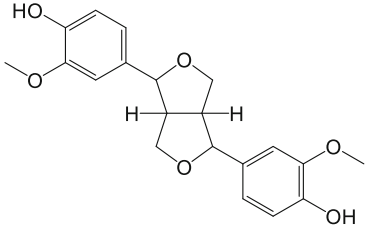
Table 2 (continued)

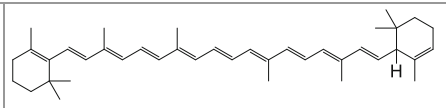
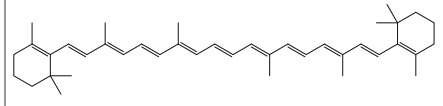
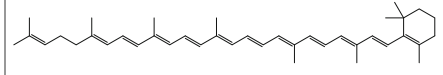
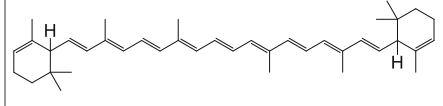
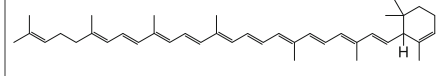
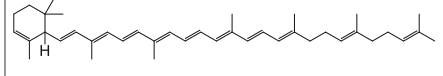
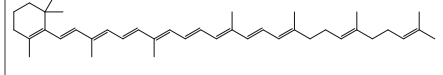
Class	Structure	Biosynthesis/pathway	References for oil palm
Phosphatidic acid (PA)	X = H		
Phosphatidylcholine (PC)	X = choline		
Phosphatidylethanolamine (PE)	X = ethanolamine		
Phosphatidylserine (PS)	X = serine		
Phosphatidylglycerol (PG)	X = glycerol		
Phosphatidylinositol (PI)	X = inositol		
Galactolipids			
Sulfoquinovosyl diacylglycerols (SQDG)		Glycerolipid metabolism (KEGG pathway map00561)	Cheong et al. (2014)
	R = fatty acid		

Terpene			
		Biosynthesis of terpenoids and steroids (KEGG pathway map01062)	Gonzalez-Diaz and García-Núñez (2021); Hoe et al. (2020)
	$(C_5H_8)_n$		
Monoterpene	$n = 2$, $C_{10}H_{16}$		
Sesquiterpene	$n =$, $C_{15}H_{24}$		
Diterpene	$n = 4$, $C_{20}H_{32}$		
Sesterpene	$n = 5$, $C_{25}H_{40}$		
Triterpene, e.g., squalene	$n = 6$, $C_{30}H_{48}$ usually contains 30 carbon atoms consisting of 6 isoprene units		
Sesquarterpene	$n = 7$, $C_{35}H_{56}$		
Tetraterpenes	$n = 8$, $C_{40}H_{64}$		
Meroterpene	Partial terpenoid skeleton		
Polyterpenes	Long chains of many isoprene units		
Terpenoid	Terpenes with additional functional groups, usually containing oxygen		

(continued)

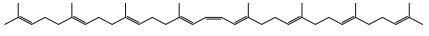
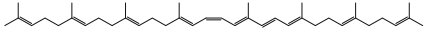
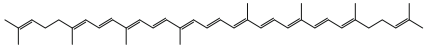
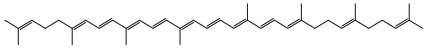
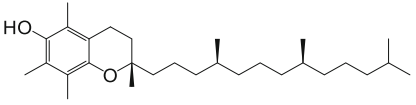
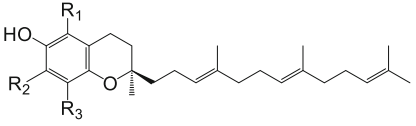
Table 2 (continued)

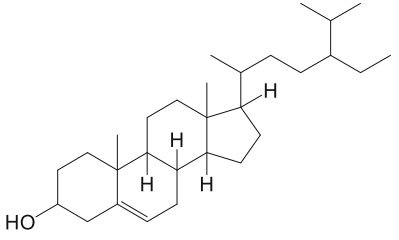
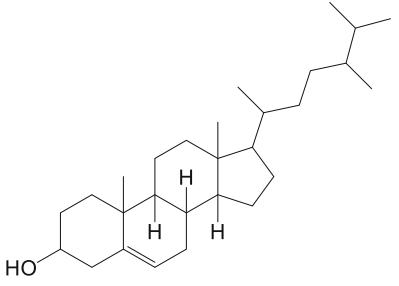
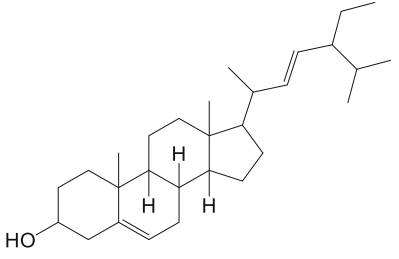
Class	Structure	Biosynthesis/pathway	References for oil palm
Lignans			
Isolariciresinol		Biosynthesis of various plant secondary metabolites (KEGG pathway map00999)	Tardugno et al. (2022)
Lariciresinol			
Pinoresinol			

Red palm oil			
Carotenoids			
α-Carotene (alpha)		Carotenoid biosynthesis pathway (KEGG pathway map00906)	Ping and Gwendoline (2006); Ng and Choo (2016)
β-Carotene (beta)			
γ-Carotene (gamma)			
ξ-Carotene (epsilon)			
δ-Carotene (delta)			
α-Zeacarotene			
β-Zeacarotene			

(continued)

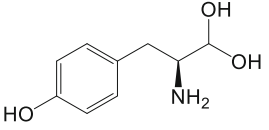
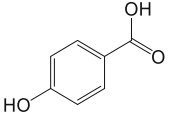
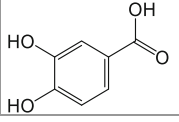
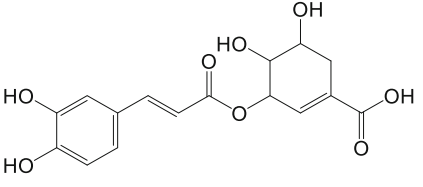
Table 2 (continued)

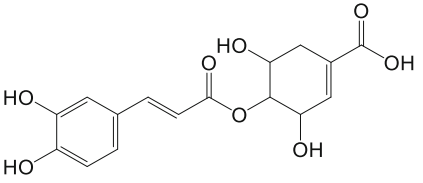
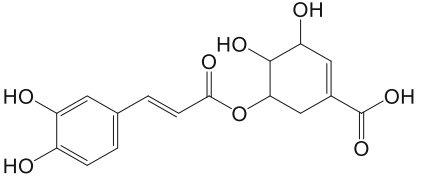
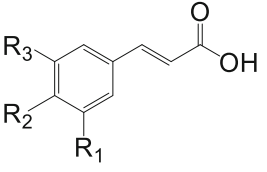
Class	Structure	Biosynthesis/pathway	References for oil palm
Phytoene			
Phytofluene			
Lycopene (ψ-carotene, psi)			
Neurosporene			
Tocols			
Tocopherol		Ubiquinone and other terpenoid-quinone biosynthesis (KEGG pathway map00130)	Hoe et al. (2020); Ibrahim (2013)
α-Tocopherol (α-T)			
Tocotrienol			
α-Tocotrienol (α-T3)	$R_1 = CH_3$, $R_2 = CH_3$, $R_3 = CH_3$		
γ-Tocotrienol (γ-T3)	$R_1 = H$, $R_2 = CH_3$, $R_3 = CH_3$		
δ-Tocotrienol (δ-T3)	$R_1 = H$, $R_2 = H$, $R_3 = CH_3$		

Phytosterols			
β-Sitosterol		Steroid biosynthesis (KEGG pathway map00100)	Jalani et al. (2021); Choo and Nesaretnam (2014)
Campesterol			
Stigmasterol			

(continued)

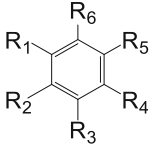
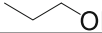
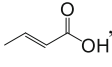
Table 2 (continued)

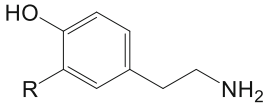
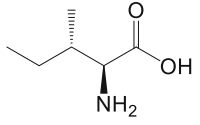
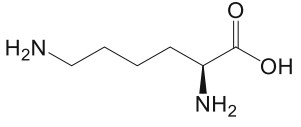
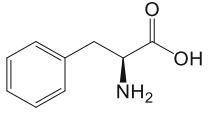
Class	Structure	Biosynthesis/pathway	References for oil palm
Palm oil mill effluent (POME)			
Phenolic acids			
Hydroxytyrosol		Tyrosine metabolism (KEGG pathway map00350)	Syarifah-Noratiqah et al. (2019) Sambanthamurthi et al. (2011)
p-Hydroxybenzoic acid		Biosynthesis of phenylpropanoids (KEGG pathway map01061)	
Protocatechuic acid		Benzoate degradation (KEGG pathway map00362)	
Isomers of caffeoylshikimic acid (CSA)			
3-*O*-caffeoylshikimic acid		Phenylpropanoid biosynthesis (KEGG pathway map00940)	

4-*O*-caffeoylshikimic acid			
5-*O*-caffeoylshikimic acid			
Hydroxycinnamic acids			
Ferulic acid	$R_1 = OCH_3$, $R_2 = OH$, $R_3 = H$		
Sinapic acid	$R_1 = OCH_3$, $R_2 = OH$, $R_3 = OCH_3$		
p-Coumaric acid	$R_1 = H$, $R_2 = OH$, $R_3 = H$		
Caffeic acid	$R_1 = OH$, $R_2 = OH$, $R_3 = H$		

(continued)

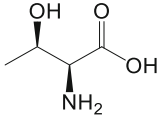
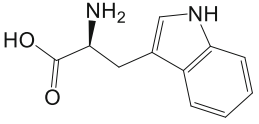
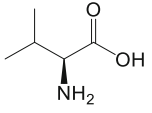
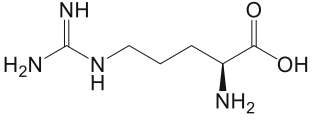
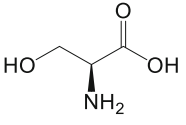
Table 2 (continued)

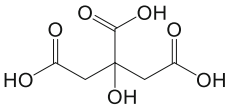
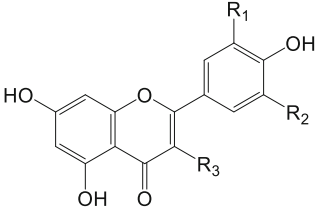
Class	Structure	Biosynthesis/pathway	References for oil palm
Phenolics (in dry matter)			
		Aminobenzoate degradation (KEGG pathway map00627)	Tsouko et al. (2019)
Gallic acid	$R_1 = OH$, $R_2 = OH$, $R_3 = OH$, $R_4 = H$, $R_5 = COOH$, $R_6 = H$		
Homovanillic alcohol	$R_1 = OH$, $R_2 = OCH_3$, $R_3 = H$, $R_4 =$  , $R_5 = H$, $R_6 = H$		
Vanillin	$R_1 = OH$, $R_2 = OCH_3$, $R_3 = H$, $R_4 = C = O$, $R_5 = H$, $R_6 = H$		
Pyrogallol	$R_1 = H$, $R_2 = H$, $R_3 = H$, $R_4 = OH$, $R_5 = OH$, $R_6 = OH$		
Catechol	$R_1 = H$, $R_2 = H$, $R_3 = H$, $R_4 = OH$, $R_5 = OH$, $R_6 = H$		
Guaiacol	$R_1 = H$, $R_2 = H$, $R_3 = H$, $R_4 = OCH_3$, $R_5 = OH$, $R_6 = H$		
Syringaldehyde	$R_1 = OH$, $R_2 = OCH_3$, $R_3 = H$, $R_4 = C = O$, $R_5 = H$, $R_6 = OCH_3$	Phenylpropanoid biosynthesis (KEGG pathway map00940)	
Sinapinic acid	$R_1 = OCH_3$, $R_2 = OH$, $R_3 = OCH_3$, $R_4 = H$, $R_5 =$  , $R_6 = H$		

Other oil palm tissues			
Amines			
β-Phenylethylamine derivatives		Biosynthesis of alkaloids derived from shikimate pathway (KEGG map01063)	Tahir et al. (2022)
Tyramine	R = H	Tyrosine metabolism (KEGG pathway map00350)	
Dopamine	R = OH		
Essential amino acid			
Isoleucine		Biosynthesis of amino acids (KEGG pathway map01230)	Rozali et al. (2021)
Lysine			
Phenylalanine			

(continued)

Table 2 (continued)

Class	Structure	Biosynthesis/pathway	References for oil palm
Threonine			
Tryptophan			
Valine			
Arginine			
Serine			

Organic acid			
Citric acid		Alanine, aspartate, and glutamate metabolism (KEGG pathway map00250)	
Flavonoids			
Flavone		Flavonoid biosynthesis (KEGG pathway map00941)	
Apigenin derivatives	$R_1 = H, R_2 = H, R_3 = H$		Tahir et al. (2012)
Luteolin derivatives	$R_1 = OH, R_2 = H, R_3 = H$		
Myricetin	$R_1 = OH, R_2 = OH, R_3 = OH$		Tsouko et al. (2019)
Flavanol			Che Zain et al. (2020)
Catechin	$R_1 = H, R_2 = OH, R_3 = OH$ (in *trans* configuration)		
Epicatechin	$R_1 = H, R_2 = OH, R_3 = OH$ (in *cis* configuration)		

4 Significance of Palm Oil and Its Phytonutrients in Human Diseases and Health

Palm oil is semisolid in a tropical environment and can be fractionated to separate the liquid component, known as palm olein, from its solid component called palm stearin. Palm olein can be used directly as cooking oil, whereas palm stearin is employed to produce fats such as margarine, shortening, and ghee. Unlike other vegetable oils that require the hydrogenation step to make them solid, more stable, and less reactive for food manufacturing, palm olein can be made more solid by adding the natural palm stearin, resulting in oils that are trans fat-free. The hydrogenation process converts the geometrically cis ("same side") structure of fatty acid molecules into the trans ("opposite side") which causes changes in physical and biological properties (Goh 2006). Consuming food that contains trans fats results in increased low-density lipoprotein (LDL), or "bad" cholesterol, and decreased "good" cholesterol (high-density lipoprotein – HDL) in blood serum and interference of many cellular signaling pathways in humans (Mensink and Katan 1990; Judd et al. 1994). Elevated levels of LDL contribute to accumulation of plaque along artery walls, also known as atherosclerosis. This subsequently causes inflammation and increases the risk of heart disease. Long-term consumption of foods blended with hydrogenated vegetable oils has also been reported to cause other major diseases such as diabetes, cancer, cystic fibrosis, diabetes, inflammatory diseases, macular degeneration, and Parkinson's and Alzheimer's diseases (Goh 2006). Studies have shown that replacing hydrogenated oils with palm olein could decrease heart disease risk by 30% (Mozaffarian and Clarke 2009). Palm olein also demonstrates anticarcinogenic and antioxidant attributes and a wide range of protective properties against diseases (Abdullah et al. 2018). Palm olein appears to be a healthier and better alternative for hydrogenated oils in the food industry as far as health effects are concerned.

RPO is trans- and cholesterol-free and has pro-vitamin A activity. Children fed with red palm oil and confectionaries containing red palm oil recorded increased blood retinol levels (Manorama et al. 1996; van Stuijvenberg et al. 2000), which significantly reduced night blindness and Bitot's spots occurrences due to vitamin A deficiency (Sivan et al. 2002). Several other studies also supported the safety and efficacy of using RPO for fighting vitamin A deficiency (Hedrén et al. 2002, Zagré et al. 2003, Canfield et al. 2001, Radhika et al. 2003,Lietz et al. 2001). RPO also potentially confers anticancer properties (Boateng et al. 2006; Yamanushi et al. 2001) and plays an important role in tumor suppression by increasing natural killer cells and B-lymphocyte populations which enhances the immune system (Nesaretnam et al. 2002).

In general, phytonutrients exhibit bioactive properties, i.e., inhibition or initiation of gene expression and enzyme activity and suppression of receptors. Phytonutrients are directly beneficial to humans as a nutraceutical or supplemental product or indirectly through use as a livestock feed (Tahir et al. 2013). Tocotrienol is the major vitamin E component extracted from palm oil. In fact, palm oil is one of the plant species offering the highest source of tocotrienols, apart from rice bran. Several

in vitro and animal studies showed that palm oil tocotrienols have anticarcinogenic potential (Komiyama et al. 1989; Sundram et al. 1989; Wada 2009; Samant et al. 2010; Shah et al. 2003; Sylvester and Shah 2005; Nesaretnam et al. 2004) and neurodegeneration protection (Nesaretnam et al. 2004; Shah et al. 2003; Sen et al. 2007; Khanna et al. 2005; Park et al. 2011). Tocotrienols could also prevent heart disease through their ability to reverse arterial blockage (Nafeeza et al. 2001; Black et al. 2000) and are involved in other cardioprotective mechanisms (Li et al. 2010; Budin et al. 2009; Das et al. 2005).

Phenolics are another phytonutrient extracted from the vegetative liquor of palm oil mill (Sambanthamurthi et al. 2011). Based on animal studies, oil palm phenolics (OPP) are beneficial for heart health and neurons (brain) and are bioactive against free radical damage and atherosclerosis (Leow et al. 2013; Sambanthamurthi et al. 2011). Palm oil and palm products with phenolics and phytosterols are strong inhibitors of peroxidation, and the dietary intake of lignan-rich foods such as oil palm products is known to prevent certain types of cancers and cardiovascular diseases. Figure 1 illustrates the human health-promoting properties of palm oil and oil palm products. Other phytonutrients such as phytosterols, squalene,

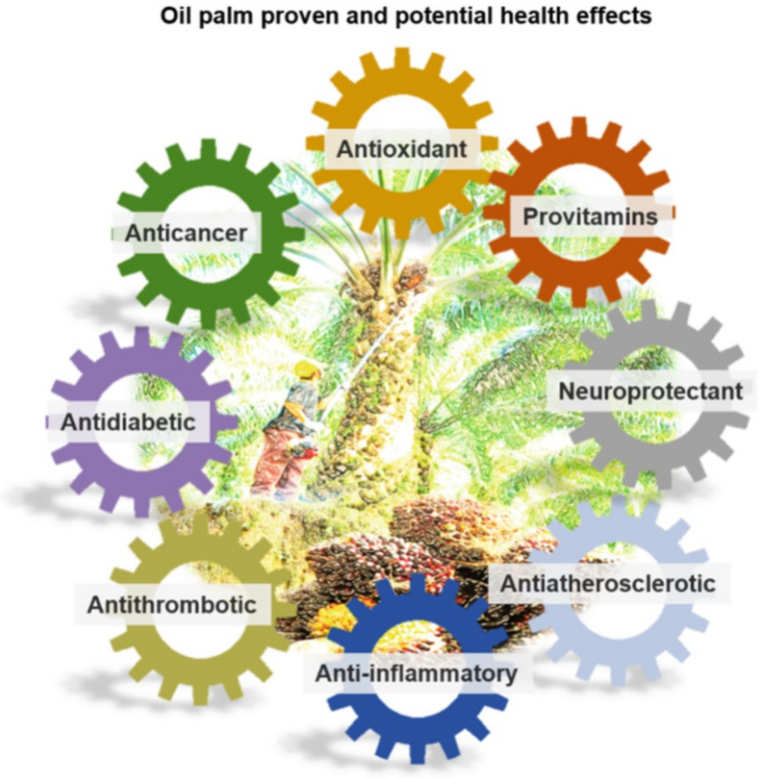


Fig. 1 Oil palm and its health-promoting properties

coenzyme Q10, phospholipids, and polyphenols play a crucial role in the stability and quality of the oil. In addition, these phytonutrients have antioxidant properties, and some exhibit nutritional and health benefits beyond their antioxidant functions (Choo and Nesaretnam 2014).

Oil palm leaves contain bioactive agents such as antioxidants and are anti-hyperglycemic with organ-protective effects against hypertension (Tahir et al. 2022). The leaf extracts also possess pharmacological activities of wound healing, hypoglycemic, vascular relaxation, hypocholesterolemic, neurogenesis, phytoestrogenic, osteogenic, fungicidal, and antimicrobial properties (Tow et al. 2021).

5 Oil Palm Genetic Diversity

The assessment of genetic diversity in any crop species helps to identify genotypes for crossing, generating hybrids with increased heterosis and variability (Adon et al. 1998). In oil palm, the diversity of germplasm populations is determined to evaluate their value for increasing diversity of the present genetically narrow breeding populations as well as identifying accessions carrying economically important genes or traits. Such efforts would benefit the oil palm industry by developing new planting materials that produce diversified products and are more sustainable against extreme climate changes as well as fatal diseases.

The genetic diversity among oil palm populations and genotypes has been evaluated by means of phenotypes and molecular markers. Phenotypic characteristics have conventionally been used to select individual plants for genetic improvement programs. Phenotypic evaluation of oil palm populations for improvement programs in Malaysia started in the 1920s. For selection and incorporation into breeding programs, oil palm breeders initially focused on evaluating oil yield and bunch quality components contributing to yield. Thus, parameters such as fresh fruit bunch (FFB), oil-to-bunch (OTB), mesocarp-to-fruit (MTF), and kernel-to-fruit (KTF) ratios are emphasized in selecting oil palms to be included in crossing programs. However, despite good progress in improving oil yield via conventional breeding, the average yield in commercial fields is 3.7 tons per hectare per year which is far below the potential yield of about 8 tons per hectare per year (Woittiez et al. 2017).

The Malaysian Palm Oil Board houses the largest oil palm genetic resource, covering both *Elaeis guineensis* and *E. oleifera*. These resources, which were collected from the species' center of distribution, are maintained as *ex situ* living collections. Field evaluation of the oil palm germplasm identified oil palms exhibiting other traits contributing to oil yield, namely, large fruit and thin-shelled (Kushairi et al. 2003) and high bunch index (Junaidah et al. 2004). Selected oil palms originating from Nigeria exhibited high oil yield (Rajanaidu et al. 2006). Incorporating these palms into breeding schemes resulted in offsprings with higher FFB, OTB, and oil yield than commercial planting materials (Marhalil et al. 2013). Besides yield, further evaluation of the oil palm germplasm populations revealed other secondary traits such as slow height increment (Rajanaidu et al. 2006), high carotene content (Mohd Din et al. 2002; Mohd Din et al. 2006), high iodine value (IV), high kernel content (Rajanaidu et al. 2006), longer stalk

(Noh et al. 2005), and high vitamin E (Kushairi et al. 2011) which gained interest among breeders. Several *Elaeis guineensis* palms, which originated from Nigeria, Cameroon, Zaire, Angola, and Tanzania, recorded high vitamin E content of between 1300 and 2496 ppm as compared to commercial varieties (800 ppm) (Kushairi et al. 2011). Carotene content ranging from 2000 to 3000 ppm, higher than the 500–700 ppm from commercial planting materials, was recorded in *E. guineensis* from Tanzania and *E. oleifera* palms from Panama and Costa Rica (Mohd Din et al. 2002, 2006). The selected palms are presently being selfed and intercrossed for introgression into oil palm genetic improvement programs to develop planting materials that produce palm oil with increased nutritional value. Apart from the simple and straightforward method of introgressing new traits from *E. guineensis*, oil palm breeders also adopted the inter-specific hybrid breeding program to integrate the fascinating traits observed in the pure *E. oleifera* into the commercial oil palm variety. This is due to the extremely low yield attained in *E. oleifera*; thus, direct commercialization of this species is not possible.

Besides phenotypic measurements, oil palm populations have been characterized using molecular marker tools. The randomly amplified polymorphic DNA (RAPD) and restriction fragment length polymorphism (RFLP) markers were initially used to study genetic diversity among Nigeria, Tanzania, Cameroon, and Zaire oil palm accessions (Shah et al. 1994; Mayes et al. 2000). Furthermore, 48 parental populations were analyzed for genetic diversity using amplified fragment length polymorphism (AFLP) markers (Purba et al. 2000). Several genetic diversity studies of *E. guineensis* and *E. oleifera* accessions originating from different collection sites were reported for pre-breeding programs using RAPD, RFLP, and AFLP markers (Moretzsohn et al. 2002; Ithnin et al. 2006). Bakoumé et al. (2015) analyzed 494 accessions representing 49 oil palm populations using simple sequence repeat (SSR) markers. This study revealed an average genetic distance value of 0.796 among accessions, indicating variability for future exploitation. The presence of rare alleles in populations from countries located within low rainfall and dry weather regions revealed their possible association with adaptive traits. The development of next-generation sequencing (NGS) techniques and expressed sequence tag-based SSRs (EST-SSRs or eSSRs) allow for assessment of the diversity of candidate genes of interest. A total of 19,243 EST sequences were mined from an oil palm EST database (http://palmoilis.mpob.gov.my/palmgenes.html) (Ting et al. 2010). Of these, 722 SSRs were designed, and polymorphisms were observed in MADS-box transcription factors, bZIP zinc finger proteins, and NAC-like transcription factors. Many more reports on applying molecular markers to analyze the genetic diversity of oil palm were recently published (Babu et al. 2019a, b, c, 2020; Bhagya et al. 2020; Sowmya et al. 2017; Venu et al. 2018; Ithnin et al. 2017; Ithnin et al. 2021).

6 Linkage and GWAS Mapping of Oil Palm

In oil palm, the relatively long generation times have promoted interest in MAS to accelerate conventional breeding through early selection. Marker trait association for oil palm was initially conducted by exploiting the genetic mapping approach

whereby a RFLP marker associated with shell thickness was identified (Mayes et al. 1997). Rance et al. (2001) and later Billotte et al. (2005) reported evidence on the linkage between SSR markers and quantitative trait loci (QTLs) for FFB yield, OTB, bunch number, oil-to-mesocarp (OTM), and vegetative components. *E. oleifera* exhibited high polyunsaturated fatty acids (PUFAs) and iodine value (IV) but low SFA compared to *E. guineensis*. Using interspecific OxG mapping populations genotyped with SSR, AFLP, and RFLP markers, Singh et al. (2009) identified QTLs influencing oleic acid, linoleic acid, palmitic acid, and iodine value. Montoya et al. (2013) studied the candidate gene mechanism involved in fatty acyl thioesterase A (FATA) and stearoyl-ACP desaturase (SAD) genes of the oleic acid pathway. Fatty acid candidate genes such as β-ketoacyl-ACP synthases (KAS) I and II were discovered within their associated QTL region on linkage group 15. It was anticipated that other candidate genes influencing other related fatty acids could also be found in the same genomic region. Pootakham et al. (2015) constructed a genetic map from 1086 single-nucleotide polymorphism (SNP) markers (Fig. 2) and identified several QTLs associated with height and fruit bunch weight in linkage groups 3, 14, and 15. Another group reported a separate QTL on a different linkage group linked to oil palm stem height (Lee et al. 2015). Palm oil quality is determined by its free fatty acid (FFA) content which is used as its quality index (Tan et al. 2009). High FFA indicates poor palm oil quality. The levels of FFAs are influenced by endogenous lipase activity which is initiated after the oil palm bunches are harvested. High lipase activity results in accumulation of FFAs and causes rancidity of palm oil. Domonhédo et al. (2018) identified the genetic loci influencing lipase activity in oil palm. This major genetic loci for lipase explained 84–92% of the variation for endogenous lipase content in oil palm. They also validated its acidification by the most common gene, viz., FLL1 (fruit lipase-like 1) and two lipase genes.

Other than genetic mapping, researchers also applied the association mapping (AM) procedure to identify loci influencing oil palm economic traits. Teh et al. (2016) employed a 200,000 SNP array to genotype over 2,000 palms and identified significant QTLs on chromosome 5 that are linked to oil-to-dry-mesocarp (OTDM) ratio, one of the key components of yield trait. It was shown that the progeny exhibiting the homozygous favorable allele for the significant SNP marker had an elevated amount of OTDM. Using the same SNP array across 312 palms, Kwong et al. (2016) detected several SNP markers on linkage groups 4, 12, and 15 influencing the shell-to-fruit (STF) ratio, also an important component influencing mesocarp content and subsequently oil content. Using over a million SNPs, Xia et al. (2019) discovered 26 significant SNPs associated with high palmitic acid and major candidate genes (acyl-ACP thioesterase B genes) within the QTL region in oil palm. Likewise, several SNP- and SSR-based molecular markers were associated with oil yield, vegetative traits, height increment, and bunch quality parameters in oil palm (Ithnin et al. 2017; Babu et al. 2019b, c, 2020; Bhagya et al. 2020; Bai et al. 2017). Bai et al. (2017) performed genotyping by sequencing across 153 oil palm DxP population and identified significant QTLs for OTB and OTDM ratios on linkage groups 1, 8, and 10. These associations were detected based on *E. guineensis*

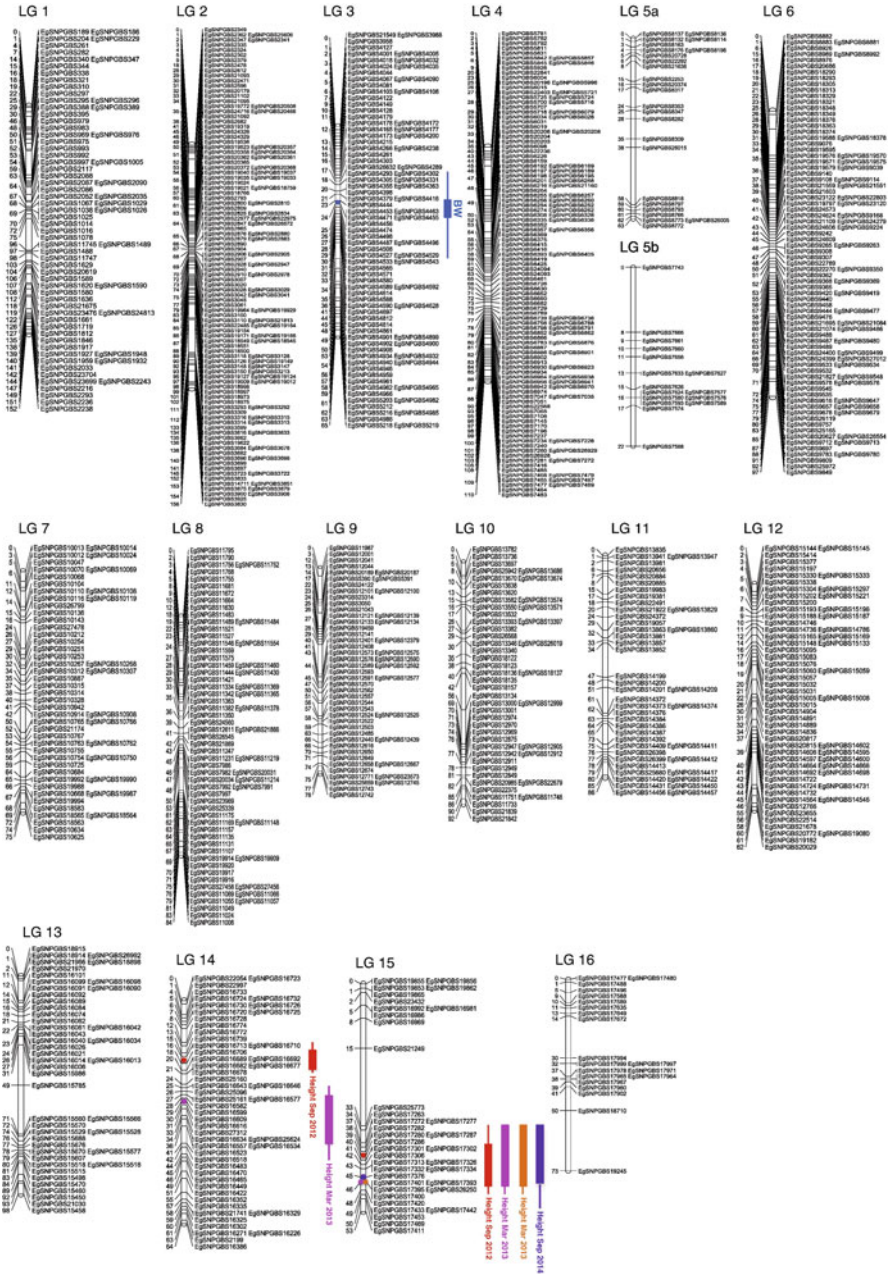


Fig. 2 Oil palm linkage map constructed using SNP markers published by Pootakham et al. (2015). QTLs associated with height and bunch weight (BW) are detected on linkage groups 3, 14, and 15

populations. Ithnin et al. (2021) described QTLs associated with fatty acid composition together with vegetative and yield-related traits from an analysis involving *E. oleifera* natural populations. The group also identified several pleiotropic SNPs linked to multiple correlated traits suggesting that the traits concerned are influenced by common genes and/or involved in the same biological pathways. Similarly, QTLs influencing fatty acid composition were detected in an interspecific (OxG) hybrid population (Shin et al. 2021). In this research, machine learning algorithms were applied to predict the values of the associated traits when different combinations of significant QTLs are used in breeding and selection programs. These outcomes provided the basis for developing MAS tools applicable for breeding programs to improve oil palm productivity and quality.

7 Oil Palm Gene and Genome Databases

Improvements in sequencing technologies and bioinformatic analyses, together with the reduced costs, have contributed to the increased number of crop genomes being sequenced each year. Although many published genomes are incomplete with variable quality, the sequences are a remarkable resource that can be exploited in various ways. The genome sequences have been used to generate high-density molecular markers to map agronomically important traits and subsequently identify candidate genes within a genomic region of interest. These traits can be characterized and bred into commercial varieties. A high-density genetic map also helps in the identification of molecular markers tightly linked to the traits of interest for MAS development. Genome sequences can be annotated with gene models to depict exons, introns, gene names, gene functions, gene regulatory sequences, and protein products, among others, allowing further research toward understanding the genetic control of gene or traits of interest. These inspired oil palm researchers to sequence the oil palm genome (Singh et al. 2013a). The oil palm genome sequence was assembled, published, and stored in Genomsawit (http://genomsawit.mpob.gov.my) (Singh et al. 2013a; Rosli et al. 2014). The database also houses the transcriptome and hypomethylated sequences useful for gene prediction and curation (Amiruddin et al. 2022; Masura et al. 2022; Amiruddin et al. 2020; Rosli et al. 2018a; Rosli et al. 2018b; Chan et al. 2017; Singh et al. 2014; Jayanthi et al. 2013; Singh et al. 2013b) and development of molecular markers (Ting et al. 2014, 2016, 2021; Yaakub et al. 2020). The transcriptome sequences are useful for the identification and analysis of expressed genes (mRNAs). Transcriptome sequencing has enabled researchers to identify differentially expressed genes between tissues, time points, and individuals with contrasting phenotypes. Transcriptomic analyses also play a part in discovering the interactions between biochemical pathways.

The Genomsawit database was integrated into a web-based interface known as MYPalmViewer (http://gbrowse.mpob.gov.my) (Low et al. 2015) for comprehensive search and interactive browsing of the data, to enable direct visualization and exploration of the oil palm genome. A total of 15 tracks were mapped to the EG5.1 genome build including oil palm genomic sequences, GenBank sequences, predicted

genes, genomic markers, other plant genomes, GC content, scaffold gaps, and enzyme restriction sites. Several biological databases were further established to store the results from advanced analysis of oil palm genomic sequences. One of them is PalmXplore-DB which consists of 26,059 EG5.1 predicted genes and their annotation (Sanusi et al. 2018). PalmXplore-DB (http://palmxplore.mpob.gov.my) is useful for the identification of new genes and gene families for biological research in oil palm. Coupled with the transcriptome sequences, research can then be initiated to further understand the biological and molecular role of individual genes and their involvement in regulatory networks. The association of gene expression data from a variety of tissues and developmental stages with gene models creates a new layer of information that contributes to the understanding of gene interactions and regulation. These are essential particularly in developing strategies using genetic engineering technology for oil palm improvement.

From the EG5.1 build, a total of 1714 SSR markers were mined and stored in the Oil Palm SSR Resource Interface (OPSRI-DB), where the primer sequences and their genome locations are available that would help link to phenotypes of interest. The OPSRI (http://opsri.mpob.gov.my) enables browsing of SSR markers together with bioinformatic tools such as Open Reading Frame Search, SSR Search, and BLAST (Rosli et al. 2022). The number of databases from the EG5.1 genome build advance analysis will be expanded in the nearest future.

The information available in Genomsawit could also facilitate researchers in their efforts to improve palm oil nutrition through marker-assisted conventional breeding techniques. Karim et al. (2021) observed an SNP conversion from TAAT to form the CAAT box which could be a promoter activity-enhancing factor leading to an increase in tocotrienol content in palm oil. Using the genomic information in Genomsawit, candidate genes associated with tocotrienol can be identified within the genomic region of the significant SNPs, for instance. New markers can be developed using the candidate gene sequences and then applied to further saturate the associated QTL regions. This will result in the identification of molecular markers tightly linked with tocotrienol in oil palm. The tightly linked QTLs can be applied in MAS to increase selection efficiency and reduce the time needed to produce oil palm with improved tocotrienol content.

8 Biofortification of Oil Palm and Palm Oil

Biofortification is the process of enhancing the nutritional quality of food crops through transgenic, conventional, and agronomic approaches, which involve modern biotechnology techniques, crop breeding, and fertilization, respectively (Garg et al. 2018). In view of the dependency of human nutritional health on plant food and agricultural products, developing different crop varieties with improved nutritional quality is of great importance. Most of the staple crops and some legumes, oilseed crops, vegetables, and fruits, such as rice, wheat, maize, barley, sorghum, soybean, pea, chickpea, canola, mustard, carrot, sweet potato, potato, lettuce, and tomato, have been nutritionally enhanced using micronutrient biofortification through

agronomic practices. Biofortification of crops by foliar spray of iron, zinc, and selenium and fortifying germinating rice plantlets with ferrous sulfate were among the effective ways to promote micronutrient concentrations in plant grains (He et al. 2013; Yuan et al. 2013; Zhang et al. 2013; Poblaciones et al. 2013). A foliar spray of selenium and iodine was favored as the most dynamic and economical method in cereals compared to soil application of the microminerals (Lyons 2018). Other than chemical and organic fertilizers, biofertilizers containing plant growth-promoting soil microorganisms, such as *Bacillus*, *Pseudomonas*, *Rhizobium*, and *Azotobacter* species, have also been applied for crop biofortification (Nooria et al. 2014; Ramzani et al. 2016; Dhawi et al. 2015; Sathya et al. 2013; Nosheen et al. 2011; Hameed et al. 2018). However, these methods are less cost-effective, labor-intensive, and non-sustainable in some cases and therefore cannot be generally applied as a strategy to improve the nutritional quality of most crops. In addition, it is not always feasible to target biofortification of micronutrient into edible plant parts. Thus, the desired nutrients may accumulate in leaves or other non-edible portions of the plants.

Crop biofortification through conventional breeding is the most preferred method on account of its success rate, sustainability, and cost-effectiveness. This approach is useful to enhance and improve the levels of micronutrient content in crops when sufficient diversity of the trait of interest is available in the gene pool of the targeted crop (Garcia-Oliveira et al. 2018). In the scenario where genetic diversity is limited or unavailable for the targeted component, genetic engineering or transformation provides another alternative for developing biofortified crops with the desired nutritional composition (Bouis and Saltzman 2017).

The improvement of oilseed nutritional quality has been targeted in some oilseed crops including oil palm, rapeseed, soybean, and sunflower (Kishore and Shewmaker 1999). Apart from increased oil palm productivity and palm oil yield, efforts are also being made to develop a nutritionally enhanced high-yielding biofortified oil palm. The improvement of fat quality traits in palm oil can be executed through biofortification of fatty acids in palm oil, where the SFA level can be controlled to an optimum level while increasing the level of bioactive components such as carotenoids and tocotrienols. Hence, more value-added palm oil can be generated in new emerging markets. This can be feasibly accomplished by identifying candidate genes and transcription factors affecting iodine value and fatty acid composition in palm oil (Singh et al. 2009). This type of oils can lead to wider dietary acceptance and weaken the link between cardiovascular disease risk and palm oil.

9 Oil Palm Genetic Engineering

Genetic engineering involves modifying and manipulating the genome to change the characteristics of an organism. It is believed to be more specific and could accelerate conventional breeding methods. Genetic engineering technique was applied to achieve sustainability in oil palm growth and development particularly improving palm oil quality, productivity, and tolerance to pests and diseases. Successful transformation strategies and regeneration methods play significant roles in the success of genetic

engineering in any crop including oil palm. The oil palm genetic engineering program started more than two decades ago, much later than other oil-bearing plants, mainly targeting to modify the fatty acid composition in the oil, specifically to increase oleic content for industrial purposes. The objectives were expanded to achieve transgenic palms that can produce palm oil with high stearate, palmitoleic, and ricinoleic acids and high lycopene as well as other products including biodegradable plastics (Parveez et al. 2015; Sambanthamurthi et al. 2002). Three methods to deliver genes into oil palm tissues have been reported, namely, particle bombardment (Parveez and Christou 1998), *Agrobacterium*-mediated transformation (Masli et al. 2009), and novel DNA microinjection (Masani et al. 2014).

Modifying the fatty acid composition in palm oil requires a change in the regulation of the genes involved in oil palm fatty acid synthesis. Some of the key genes concerned have been isolated and characterized, including acetyl-CoA carboxylase (Omar et al. 2008), 3-keto-acyl-ACP synthase II (KAS-II) (Ramli and Sambanthamurthi 1996), palmitoyl-acyl carrier protein thioesterase (Abrizah et al. 1999; Parveez et al. 2010), stearoyl-ACP desaturase (Shah et al. 2000), oleoyl-ACP thioesterase (Asemota et al. 2004), oleoyl-ACP desaturase (Syahanim et al. 2007), and lysophosphatidic acid acyltransferase (Manaf et al. 2005). In addition, a handful of genes involved in the carotenoid synthesis including phytoene synthase (Rasid et al. 2008), phytoene desaturase (Rasid et al. 2014), lycopene cyclases (Rasid et al. 2009), and 1-deoxy-D-xylulose-5-phosphate synthase (Khemvong and Suvachittanont 2005) have also been isolated. Analysis of gene expression using RT-PCR revealed that phytoene desaturase and lycopene cyclase convert lycopene to β-carotene (Römer et al. 2000). Thus, genes for both enzymes are important for the genetic engineering of oil palm targeting palm oil with higher β-carotene content.

Regulatory sequences, also known as promoters, which are capable of increasing or decreasing the expression of specific genes in various oil palm tissues, are available. Oil palm promoters that have been identified include the mesocarp-specific (MT3-A) (Zubaidah et al. 2017) and FLL1 (Nurniwalis et al. 2015), kernel-specific (pOP-KT21) (Siti Nor Akmar et al. 2014), root-specific (MT3-B) (Zubaidah and Abdullah 2010), and leaf-specific (LS01) (Chan et al. 2008; Hanin et al. 2016) promoters. With the availability of these genetic materials and establishment of the transformation technique, positive transformants carrying the respective genes have been produced. Palmitoyl acyl carrier protein thioesterase (PATE) that regulates the accumulation of palmitic acid in palm oil was successfully transformed into *E. oleifera* immature zygotic embryos using the bombardment transformation method (Bhore and Shah 2012). Downregulation of this gene will result in improved palm oil quality with reduced palmitic acid content.

The *Cry* genes isolated from *Bacillus thuringiensis*, which are effective in conferring insect resistance in other crops, were introduced into oil palm immature embryos via the biolistic method (Lee et al. 2006). Further analysis revealed successful expression of the transgene in the transformed tissues. In addition, the transformation method to deliver glufosinate-ammonium resistant genes into oil palm tissues using the alternative *Agrobacterium* system was reported by Masli et al. (2009). Although the transformation rate attained was lower than that achieved

using the biolistic method in oil palm, the *Agrobacterium*-mediated technique offers a stable integration of the transgene at a low copy number.

Yeap et al. (2021) recently reported their success in establishing a gene editing method, namely, clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) mutagenesis system for the oil palm. A mutation frequency between 62.5% and 83.33% was recorded when CRISPR was applied to phytoene desaturase (EgPDS), a gene involved in carotenoid biosynthesis. Mutation of this gene resulted in albino phenotypes among the transgenic oil palm shoots. Furthermore, application of CRISPR/Cas9 on EgBRI1 gene, coding for a major receptor of the plant hormone brassinosteroid, resulted in a stunted phenotype of oil palm transgenic shoots due to nucleotide substitutions. The CRISPR/Cas9 method was also employed to knock out the OsFAD2-1 gene in rice as a model system (Bahariah et al. 2021). The goal is to subsequently repeat this highly effective approach in knocking out fatty acid desaturase genes in oil palm, to produce higher oleic acid in palm oil. These achievements lay the foundation for adopting gene editing techniques for nutrition-enhancement programs in oil palm.

10 Conclusion

The enormous amounts of vitamins and phytonutrients present in palm oil have opened up substantial opportunities in research leading to significant discoveries of their health and nutritional benefits. The available germplasm collection and genomic sequences of oil palm provide great opportunities in developing new planting materials to produce palm oil with enhanced nutritional values. Using novel genomic, genetic, and molecular techniques, relevant biochemical pathways can be investigated which leads to an understanding of the genetic control mechanisms underlying the synthesis and accumulation of essential vitamins and phytonutrients. Such knowledge is useful in formulating a more efficient and rapid improvement process through MAS for conventional breeding and/or development of transgenic lines aimed at increasing vitamins and phytonutrient levels in palm oil. While improvement efforts via conventional breeding are rather straightforward, implementing genetic engineering approaches faces a complex legal framework. Nevertheless, such efforts are worthy, particularly when implemented in a highly productive crop like the oil palm. Moreover, palm oil has a wide range of applications in the food industry. Thus, improving palm oil's nutritional value could intensify efforts toward delivering more nutritious food in addressing the malnutrition issues faced globally.

References

Abd Rashid SNA, Ab Malik S, Embi K, Mohd Ropi NA, Yaakob H et al (2021) Carotenoids and antioxidant activity in virgin palm oil (VPO) produced from palm mesocarp with low heat aqueous-enzyme extraction techniques. Mater Today Proc 42:148–152. https://doi.org/10.1016/j.matpr.2020.10.616

Abdullah F, Ismail R, Ghazali R, Idris Z (2018) Total phenolic contents and antioxidant activity of palm oils and palm kernel oils at various refining processes. J Oil Palm Res 30(4):682–692. https://doi.org/10.21894/jopr.2018.0058

Abrizah O, Lazarus CM, Stobart AK (1999) Isolation of a cDNA clone encoding an acyl-acyl carrier protein thioesterase from the mesocarp of oil palm (*Elaeis guineensis*). J Oil Palm Res (Special Issue):81–87

Adon B, Baudouin L, Durand GT, Kouame B (1998) Use of non-selected material for oil palm breeding: Angolan germplasm. Plantations, Recherche, Développement 5(3):201–207

Ahsan H, Ahad A, Siddiqui WA (2015) A review of characterization of tocotrienols from plant oils and foods. J Chem Biol 20;8(2):45–59. https://doi.org/10.1007/s12154-014-0127-8

Alseekh S, Perez de Souza L, Benina M, Fernie AR (2020) The style and substance of plant flavonoid decoration; towards defining both structure and function. Phytochemistry 174: 112347. https://doi.org/10.1016/j.phytochem.2020.112347

Amiruddin N, Chan PL, Azizi N, Morris PE, Chan KL (2020) Characterization of oil palm acyl-CoA-binding proteins and correlation of their gene expression with oil synthesis. Plant Cell Physiol 61(4):735–747

Amiruddin N, Chan PL, Chan KL, Ong PW, Morris PE et al (2022) Determination of reliable reference genes for reverse transcription quantitative real-time PCR from oil palm transcriptomes. J Oil Palm Res 34(3):439–452

Asemota O, San CT, Shah FH (2004) Isolation of a kernel oleoyl-ACP thioesterase gene from the oil palm *Elaeis guineensis* Jacq. Afr J Biotechnol 3(3):199–201

Babu BK, Mary Rani KL, Sarika Sahu RK, Mathur PN, Kumar GR et al (2019a) Development and validation of whole genome-wide and genic microsatellite markers in oil palm (*Elaeis guineensis* Jacq.): first microsatellite database (OpSatdb). Sci Rep (Nature) 9(1):1899. https://doi.org/10.1038/s41598-01837737-7

Babu BK, Mathur RK, Ravichandran G, Venu MVB (2019b) Genome-wide association study (GWAS) for stem height increment in oil palm (*Elaeis guineensis*) germplasm using SNP markers. Tree Genet Genomes 15:40. https://doi.org/10.1007/s11295-019-1349-2

Babu BK, Mathur RK, Ravichandran G, Anitha P, Venu MVB (2019c) Genome-wide association study for leaf area, rachis length and total dry weight in oil palm (*Elaeis guineensis*) using genotyping by sequencing. PLoS One 14(8):e0220626. https://doi.org/10.1371/journal.pone.0220626

Babu BK, Mathur RK, Ravichandran G, Anita P, Venu MVB (2020) Genome wide association study (GWAS) and identification of candidate genes for yield and oil yield related traits in oil palm (*Elaeis guineensis*) using SNPs by genotyping-based sequencing. Genomics. https://doi.org/10.1016/j.ygeno.2019.06.018

Bahariah B, Masani MYA, Rasid OA, Parveez GKA (2021) Multiplex CRISPR/Cas9-mediated genome editing of the FAD2 gene in rice: a model genome editing system for oil palm. J Genet Engin Biotechnol 19(1):1–13

Bai B, Wang L, Lee M, Zhang Y, Rahmadsyah YA et al (2017) Genome-wide identification of markers for selecting higher oil content in oil palm. BMC Plant Biol 17:93

Bakoumé C, Wickneswari R, Siju S, Rajanaidu N, Kushairi A (2015) Genetic diversity of the world's largest oil palm (*Elaeis guineensis* Jacq.) field genebank accessions using microsatellite markers. Genet Resour Crop Evol 62(3):349–360

Bhagya HP, Kalyana B, Gangadharappa PM, Naika MB, Satish D et al (2020) Identification of QTLs in oil palm (*Elaeis guineensis* Jacq.) using SSR markers through association mapping. J Genet 99(1):1–10

Bhore SJ, Shah FH (2012) Genetic transformation of the American oil palm (*Elaeis oleifera*) immature zygotic embryos with antisense palmitoyl-acyl carrier protein thioesterase (PATE) gene. World Appl Sci J 16(3):362–369

Billotte N, Marseillac N, Risterucci AM, Adon B, Brottier P et al (2005) Microsatellite based high density linkage map in oil palm (*Elaeis guineensis* Jacq.). Theor Appl Genet 110:754–765

Black TM, Wang P, Maeda N, Coleman RA (2000) Palm tocotrienols protect ApoE +/– mice from diet-induced atheroma formation. J Nutr 130:2420–2426

Boateng J, Verghese M, Chawan CB, Shackelford L, Walker LT et al (2006) Red palm oil suppresses the formation of azoxy-methane (AOM) induced aberrant crypt foci (ACF) in Fisher 344 male rats. Food Chem Toxicol 44:1667–1673

Bouis HE, Saltzman A (2017) Improving nutrition through biofortification: a review of evidence from Harvestplus, 2003 through 2016. Glob Food Sec 12:49–58. https://doi.org/10.1016/j.gfs.2017.01.009

Budin SB, Othman F, Louis SR, Bakar MA, Das S et al (2009) The effects of palm oil tocotrienol-rich fraction supplementation on biochemical parameters, oxidative stress and the vascular wall of streptozotocin-induced diabetic rats. Clinics 64:235–244

Bustamam FKA, Yeoh CB, Sulaiman N (2019) Revision of the Malaysian standard for palm oil – specification MS 814:2007, (Second Revision) Amendment 1:2018 – What's New? Palm Oil Developments 71:18–32

Canfield LM, Kaminsky RG, Taren DL, Shaw E, Sander JK (2001) Red palm oil in the maternal diet increases provitamin A carotenoids in breastmilk and serum of the mother-infant dyad. Eur J Nutr 40(1):30–38

Cazzaniga S, Bressan M, Carbonera D, Agostini A, Dall'Osto L (2016) Differential roles of carotenes and xanthophylls in photosystem I photoprotection. Biochemist 55(26):3636–3649. https://doi.org/10.1021/acs.biochem.6b00425

Chan PL, Siti Nor Akmar A, Roohaida O (2008) Light-harvesting chlorophyll A/B binding protein (LHCB) promoter for targeting specific expression for oil palm leaves. J Oil Palm Res (Special Issue):21–29

Chan KL, Tatarinova TV, Rosli R, Amiruddin N, Azizi N et al (2017) Evidence-based gene models for structural and functional annotations of the oil palm genome. Biol Direct 12(1):1–23

Cheong WF, Wenk MR, Shui G (2014) Comprehensive analysis of lipid composition in crude palm oil using multiple lipidomic approaches. J Genet Genomes 41(5):293–304. https://doi.org/10.1016/j.jgg.2014.04.002

Che Zain MS, Lee SY, Nasir NM, Fakurazi S and Shaari K (2020) Metabolite characterization and correlations with antioxidant and wound healing properties of oil palm (Elaeis guineensis Jacq.) leaflets via 1H-NMR-based metabolomics approach. Molecules 25(23):5636. https://doi.org/10.3390/molecules25235636

Choo YM, Nesaretnam K (2014) Research advancements in palm oil nutrition. Eur J Lipid Sci Technol 116:1301–1315. https://doi.org/10.1002/ejlt.201400076

Choo YM, Bong SC, Ma AN, Chuah CH (2004) Phospholipids from palm-pressed fiber. J Am Oil Chem Soc 81(5):471–475. https://doi.org/10.1007/s11746-004-0925-4

Christensen LP (2009) Galactolipids as potential health promoting compounds in vegetable foods. Recent Pat Food Nutr Agric 1(1):50–58. https://doi.org/10.2174/2212798410901010050

Ciucure CT, Geană E (2019) Phenolic compounds profile and biochemical properties of honeys in relationship to the honey floral sources. Phytochem Anal 30(4):481–492. https://doi.org/10.1002/pca.2831

Csepregi K, Hideg É (2017) Phenolic compound diversity explored in the context of photo-oxidative stress protection. Phytochem Anal 29(2):129–136. https://doi.org/10.1002/pca.2720

Das S, Powell SR, Wang P, Divald A, Nesaretnam K et al (2005) Cardioprotection with palm tocotrienol: antioxidant activity of tocotrienol is linked with its ability to stabilize proteasomes. Am J Physiol Heart Circ Physiol 289:H361–H367

Davin LB, Lewis NG (2000) Dirigent proteins and dirigent sites explain the mystery of specificity of radical precursor coupling in lignan and lignin biosynthesis. Plant Physiol 123(2):453–462. https://doi.org/10.1104/pp.123.2.453

De Smet E, Mensink RP, Plat J (2012) Effects of plant sterols and stanols on intestinal cholesterol metabolism: suggested mechanisms from past to present. Mol Nutr Food Res 56(7):1058–1072. https://doi.org/10.1002/mnfr.201100722

Dhawi F, Datta R, Ramakrishna W (2015) Mycorrhiza and PGPB modulate maize biomass, nutrient uptake and metabolic pathways in maize grown in mining-impacted soil. Plant Physiol Biochem 97:390–399. https://doi.org/10.1016/j.plaphy.2015.10.028

Domonhédo H, Cuéllar T, Espeout S, Droc G, Summo M et al (2018) Genomic structure, QTL mapping, and molecular markers of lipase genes responsible for palm oil acidity in the oil palm (*Elaeis guineensis* Jacq.). Tree Genet Genomes 14:69. https://doi.org/10.1007/s11295-018-1284-7

Ebrahimi M, Rajion MA, Meng GY, Shokryzadan P, Sazili AQ et al (2015) Feeding oil palm (*Elaeis guineensis* Jacq.) fronds alters rumen protozoal population and ruminal fermentation pattern in goats. Ital J Anim Sci 14(3):3877. https://doi.org/10.4081/ijas.2015.3877

Garcia-Oliveira AL, Chander S, Ortiz R, Menkir A, Gedil M (2018) Genetic basis and breeding perspectives of grain iron and zinc enrichment in cereals. Front Plant Sci 9:937. https://doi.org/10.3389/fpls.2018.00937

Garg M, Sharma N, Sharma S, Kapoor P, Kumar et al (2018) Biofortified crops generated by breeding, agronomy, and transgenic approaches are improving lives of millions of people around the world. Front Nutr 5:12. https://doi.org/10.3389/fnut.2018.00012

Goh SH (2006) Oils and fats in nutrition and health: 2. Chemistry, digestion and metabolism. Malays Oil Sci Technol 15:43–63

Goh SH, Khor HT, Gee PT (1982) Phospholipids of palm oil (*Elaeis guineensis*). J Am Oil Chem Soc 59:296–299. https://doi.org/10.1007/BF02662230

Gold IL, Oseni NT, Imoisi OB (2016) Oil palm fruit calyx as a resource for phospholipids extraction. Afr J Biochem Res 10(4):25–29. https://doi.org/10.5897/AJBR2016.0884

Gonzalez-Diaz A, García-Núñez JA (2021) Minor compounds of palm oil: properties and potential applications. In: Kamyab H (ed) *Elaeis guineensis*. IntechOpen, London. https://doi.org/10.5772/intechopen.99526

Gundogdu M, Ozrenk K, Ercisli S, Kan T, Kodad O et al (2014) Organic acids, sugars, vitamin C content and some pomological characteristics of eleven hawthorn species (*Crataegus* spp.) from Turkey. Biol Res 47:21. https://doi.org/10.1186/0717-6287-47-21

Hameed A, Zaidi SS, Shakir S, Mansoor S (2018) Applications of new breeding technologies for potato improvement. Front Plant Sci 9:925. https://doi.org/10.3389/fpls.2018.00925

Hanin AN, Masani MYA, Parveez GKA (2016) Evaluation of oil palm leaf-specific promoter (LSP1) activity for expressing PHB genes in *Arabidopsis thaliana*. J Oil Palm Res 28(1):1–9

Hardon JJ (1969) Interspecific hybrids in the genus *Elaeis* II. Vegetative growth and yield of F1 hybrids *E. guineensis* x *E. oleifera*. Euphytica 18(3):380–388

Hassan NSM, Hossain MS, Balakrishnan V, Zuknik MH, Mustaner M et al (2021) Influence of fresh palm fruit sterilization in the production of carotenoid-rich virgin palm oil. Foods 10:2838. https://doi.org/10.3390/foods10112838

Hazir MHM, Shariff ARM, Amiruddin MD (2012) Determination of oil palm fresh fruit bunch ripeness based on flavonoids and anthocyanin content. Ind Crop Prod 36(1):466–475. https://doi.org/10.1016/j.indcrop.2011.10.020

He W, Shohag MJ, Wei Y, Feng Y, Yang X (2013) Iron concentration, bioavailability, and nutritional quality of polished rice affected by different forms of foliar iron fertilizer. Food Chem 141(4):4122–4126. https://doi.org/10.1016/j.foodchem.2013.07.005

Hedrén E, Mulokozi G, Svanberg U (2002) *In vitro* accessibility of carotenes from green leafy vegetables cooked with sunflower oil or red palm oil. Int J Food Sci Nutr 53(6):445–453

Hoe BC, Chan E, Nagasundara Ramanan R, Ooi CW (2020) Recent development and challenges in extraction of phytonutrients from palm oil. Comprehensive Reviews in Food Science and Food Safety 19(6):4031–4061. https://doi.org/10.1111/1541-4337.12648

Ibrahim NA (2013) Characteristics of Malaysian palm kernel and its products. J Oil Palm Res 25(2):245–252

Ithnin M, Rajanaidu N, Zakri AH, Cheah SC (2006) Assessment of genetic diversity in oil palm (*Elaeis guineensis* Jacq.) using Restriction Fragment Length Polymorphism (RFLP). Genet Resour Crop Evol 53(1):187–195

Ithnin M, Xu Y, Marjuni M, Serdari NM, Amiruddin MD et al (2017) Multiple locus genome-wide association studies for important economic traits of oil palm. Tree Genet Genomes 13(5):1–14

Ithnin M, Vu WT, Shin MG, Suryawanshi V, Sherbina K, Zolkafli SH et al (2021) Genomic diversity and genome-wide association analysis related to yield and fatty acid composition of wild American oil palm. Plant Sci 304:110731

Jalani NF, Noorshamsiana AW, Wan Hasamudin WH, Hassan K, Zainal NH et al (2021) Extraction and purification of phytosterols mixture from palm fatty acid distillate (PFAD) using multistage extraction processes. J Oil Palm Res 33(1):93–102. https://doi.org/10.21894/jopr.2020.0067

Japir AA-W, Salimon J, Derawi D, Bahadi M, Al-Shuja'a S, Yusop MR (2017) Physicochemical characteristics of high free fatty acid crude palm oil. OCL 24(5):D506. https://doi.org/10.1051/ocl/2017033

Jayanthi N, Rozana R, Ting NC, Ooi LCL, Low ETL et al (2013) Exploiting synteny between oil palm and rice to find markers more closely linked to selected trait. J Oil Palm Res 25(2): 180–187

Judd JT, Clevidence BA, Muesing RA, Wittes J, Sunkin ME et al (1994) Dietary trans fatty acids: effects on plasma lipids and lipoproteins of healthy men and women. Am J Clin Nutr 59: 861–868

Junaidah J, Kushairi AD, Isa ZA, Mohd Din A, Noh A et al (2004) PS7: high bunch index breeding population. MPOB TT No 221 Malaysian Palm Oil Board, Kuala Lumpur

Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28(1):27–30. https://doi.org/10.1093/nar/28.1.27

Karim MSN, Abdullah SNA, Nadzir MMM, Moradpour M, Shaharuddin NA et al (2021) Single nucleotide polymorphisms in oil palm homogentisate geranyl-geranyl transferase promoter for species differentiation and tocotrienol improvement. Meta Gene 27:100818

Khanna S, Roy S, Slivka A, Craft TK, Chaki S (2005) Neuroprotective properties of the natural vitamin E α-tocotrienol. Stroke 36(10):e144–e152

Khemvong S, Suvachittanont W (2005) Molecular cloning and expression of a cDNA encoding 1-deoxy-d-xylulose-5-phosphate synthase from oil palm *Elaeis guineensis* Jacq. Plant Sci 169(3):571–578

Khosla P, Sundram K (2010) A supplement on palm oil – why? J Am Coll Nutr 29:237S–239S

Kishore GM, Shewmaker C (1999) Biotechnology: enhancing human nutrition in developing and developed worlds. Proc Natl Acad Sci USA 96:5968–5972. https://www.pnas.org/doi/pdf/10.1073/pnas.96.11.5968

Komiyama S, Matsui K, Kudoh S, Nogae I, Kuratomi Y (1989) Establishment of tumor cell lines from a patient with head and neck cancer and their different sensitivities to anti-cancer agents. Cancer 63(4):675–681

Kushairi A, Rajanaidu N, Mohd Din A (2003) Mining the germplasm. In: Proceedings of ISOPB Seminar on the Progress of Oil Palm Breeding and Selection. Medan, Sumatra, Indonesia, pp 6–9

Kushairi A, Rajanaidu N, Sundram K, Maizura I (2011) PS6: high vitamin E breeding population, PORIM TT no 222. Malaysian Palm Oil Board, Kuala Lumpur

Kwong QB, Teh CK, Ong AL, Heng HY, Lee HL et al (2016) Development and validation of a high-density SNP genotyping array for African oil palm. Mol Plant 9(8):1132–1141

Lee MP, Yeun LH, Abdullah R (2006) Expression of *Bacillus thuringiensis* insecticidal protein gene in transgenic oil palm. Electron J Biotechnol 9(2):117–126

Lee M, Xia JH, Zou Z, Ye J, Alfiko Y (2015) A consensus linkage map of oil palm and a major QTL for stem height. Sci Rep 5(1):1–7

Leow KY, Goh WWB, Tan SZ, Lim J, Ng K et al (2013) Variant screening of the serum amyloid A1 gene and functional study of the p. Gly90Asp variant for its role in atherosclerosis. Atherosclerosis 227(1):112–117

Li F, Tan W, Kang Z, Wong CW (2010) Tocotrienol enriched palm oil prevents atherosclerosis through modulating the activities of peroxisome proliferators-activated receptors. Atherosclerosis 211:278–282

Lietz G, Henry CJK, Mulokozi G, Mugyabuso JK, Ballart A et al (2001) Comparison of the effects of supplemental red palm oil and sunflower oil on maternal vitamin A status. Amer J Clin Nutr 74(4):501–509

Liu RH (2004) Potential synergy of phytochemicals in cancer prevention: mechanism of action. J Nutr 134(12):3479S–3485S. https://doi.org/10.1093/jn/134.12.3479S

Loganathan R, Subramaniam KM, Radhakrishnan AK, Choo YM, Teng KT (2017) Health-promoting effects of red palm oil: evidence from animal and human studies. Nutr Rev 75(2): 98–113. https://doi.org/10.1093/nutrit/nuw054

Low ETL, Ab Halim MA, Rosli R, Chan KL, Azizi N et al (2015) MYPalmViewer: oil palm genome browser, MPOB TS 148. Malaysian Palm Oil Board, Kuala Lumpur

Lyons G (2018) Biofortification of cereals with foliar selenium and iodine could reduce hypothyroidism. Front Plant Sci 9:730. https://doi.org/10.3389/fpls.2018.00730

Manaf MAA, Abrizah O, Umi Salamah R (2005) Characterization of genes encoding key enzymes in oil synthesis in the oil palm. In: Proc of the PIPOC 2005 Int Palm Oil Congress (Agriculture, Biotechnology & Sustainability), pp 583–606

Mancini A, Imperlini E, Nigro E, Montagnese C, Daniele A, Orrù S, Buono P (2015) Biological and nutritional properties of palm oil and palmitic acid: effects on health. Molecules 18;20 (9):17339–17361. https://doi.org/10.3390/molecules200917339

Manorama R, Brahmam GNV, Rukmini C (1996) Red palm oil as a source of β-carotene for combating vitamin A deficiency. Plant Food Hum Nutr 49:75–82. https://doi.org/10.1007/BF01092524

Marhalil M, Rafii MY, Afizi MMA, Arolu IW, Noh et al (2013) Genetic variability in yield and vegetative traits in elite germplasm of MPOB-Nigerian dura x AVROS pisifera progenies. J Food Agric Environ 11:515–519

Masani MYA, Noll GA, Parveez GKA, Sambanthamurthi R, Prüfer D (2014) Efficient transformation of oil palm protoplasts by PEG-mediated transfection and DNA microinjection. PLoS One 9:e96831. https://doi.org/10.1371/journal.pone.0096831

Masli DIA, Parveez GKA, Masani MYA (2009) Transformation of oil palm using *Agrobacterium tumefaciens*. J Oil Palm Res 21:643–652

Masura SS, Abd Rasid O, Shaharuddin NA, Masani MYA, Chan KL et al (2022) Identification of oil palm root-specific genes through mining of RNA-SEQ data and RT-qPCR analysis. J Oil Palm Res 34(2):261–275

Masyita A, Sari RM, Astuti AD, Yasir B, Rumata NR et al (2022) Terpenes and terpenoids as main bioactive compounds of essential oils, their roles in human health and potential application as natural food preservatives. Food Chem 13:100217. https://doi.org/10.1016/j.fochx.2022.100217

Mayes S, Jack PL, Corley RHV, Marshall DF (1997) Construction of a RFLP genetic linkage map for oil palm (*Elaeis guineensis* Jacq.). Genome 40(1):116–122

Mayes S, Jack PL, Corley RHV (2000) The use of molecular markers to investigate the genetic structure of an oil palm breeding programme. Heredity 85(3):288–293

Mazlan NA, Kamaliah AS, Hafizuddin WY, Rozaimi AS, Jamaliah MJ (2021) Xylan recovery from dilute nitric acid pretreated oil palm frond bagasse using fractional factorial design. J Oil Palm Res 33(2):307–319. https://doi.org/10.21894/jopr.2020.0093

Mensink RP, Katan MB (1990) Effect of dietary trans fatty acids on high-density and low-density lipoprotein cholesterol levels in healthy subjects. N Engl J Med 323:439–445

Mohd Din A, Rajanaidu N, Kushairi A, Mohd Rafii Y, Mohd Isa Z et al (2002) PS4 – high carotene E. oleifera planting materials, MPOB TT no 137. Malaysian Palm Oil Board, Kuala Lumpur

Mohd Din A, Noh A, Mohd Isa ZA, Maizura I, Kushairi A et al (2006) PS11 – high carotene *E. guineensis* breeding population, MPOB TT no 312. Malaysian Palm Oil Board, Kuala Lumpur

Montoya C, Lopes R, Albert F et al (2013) Quantitative trait loci (QTLs) analysis of palm oil fatty acid composition in an interspecific pseudo-backcross from *Elaeis oleifera* (H.B.K.) Corte's and oil palm (*Elaeis guineensis* Jacq.). Tree Genet Genomes 9:1207–1225

Moretzsohn MDC, Ferreira MA, Amaral ZPS, Coelho PJDA, Grattapaglia D et al (2002) Genetic diversity of Brazilian oil palm (*Elaeis oleifera* HBK) germplasm collected in the Amazon Forest. Euphytica 124(1):35–45

Mozaffarian D, Clarke R (2009) Quantitative effects on cardiovascular risk factors and coronary heart disease risk of replacing partially hydrogenated vegetable oils with other fats and oils. Eur J Clin Nutr 63:S22–S33

Muñoz IJ, Schilman PE, Barrozo RB (2020) Impact of alkaloids in food consumption, metabolism and survival in a blood-sucking insect. Sci Rep 10:9443. https://doi.org/10.1038/s41598-020-65932-y

Nafeeza MI, Norzana AG, Jalaluddin HL, Gapor MT (2001) The effects of a tocotrienol-rich fraction on experimentally induced atherosclerosis in the aorta of rabbits. Malays J Pathol 23: 17–25

Naithani R, Huma LC, Holland LE, Shukla D, McCormick DL et al (2008) Antiviral activity of phytochemicals: a comprehensive review. Mini-Rev Med Chem 8(11):1106–1133. https://doi.org/10.2174/138955708785909943

Nesaretnam K, Radhakrishnan A, Selvaduray KR, Reimann K, Pailoor J et al (2002) Effect of palm oil carotene on breast cancer tumorigenicity in nude mice. Lipids 37:557–560

Nesaretnam K, Ambra R, Selvaduray KR, Radhakrishnan A, Reimann K et al (2004) Tocotrienol-rich fraction from palm oil affects gene expression in tumors resulting from MCF-7 cell inoculation in athymic mice. Lipids 39(5):459–467

Ng MH, Choo YM (2016) Improved method for the qualitative analyses of palm oil carotenes using UPLC. J Chromatogr Sci 54(4):633–638. https://doi.org/10.1093/chromsci/bmv241

Noh A, Kushairi A, Mohd Din A, Maizura I, Mohd Isa ZA, Rajanaidu N (2005) PS10: breeding population selected for long stalk, PORIM TT no 222. Malaysian Palm Oil Board, Kuala Lumpur

Nooria M, Adibiana M, Sobhkhizia A, Eyidozehib K (2014) Effect of phosphorus fertilizer and mycorrhiza on protein percent, dry weight, weight of 1000 grain in wheat. Int J Plant Anim Environ Sci 4(2):561–564. https://www.fortunejournals.com/ijpaes/admin/php/uplo

Nosheen A, Bano A, Ullah F (2011) Nutritive value of canola (Brassica napus L.) as affected by plant growth promoting rhizobacteria. Eur J Lipid Sci Technol 113(11):1342–1346. https://doi.org/10.1002/ejlt.201000549

Nurazah Z, Idris AS, Kushairi A, Ramli US (2013) Metabolite profiling of oil palm towards understanding basal stem rot (BSR) disease. J Oil Palm Res 25(1):58–71

Nurniwalis AW, Zubaidah R, Siti Nor Akmar A, Zulkifli H, Arif M et al (2015) Genomic structure and characterization of a lipase class 3 gene and promoter from oil palm. Biol Plant 59(2): 227–236

Ofori-Boateng C, Lee KT (2013) Sustainable utilization of oil palm wastes for bioactive phytochemicals for the benefit of the oil palm and nutraceutical industries. Phytochem Rev 12: 173–190. https://doi.org/10.1007/s11101-013-9270-z

Oil World (2022) Oil world annual 2022, 3 June 2022. ISTA Mielke GmbH, Hamburg

Omar WSW, Willis LB, Rha C, Sinskey AJ, Ramli US et al (2008) Isolation and utilization of acetyl-coa carboxylase from oil palm (*Elaeis guineensis*) mesocarp. J Oil Palm Res 2:97–107

Palamae S, Dechatiwongse P, Choorit W, Chisti Y, Prasertsan P (2017) Cellulose and hemicellulose recovery from oil palm empty fruit bunch (EFB) fibers and production of sugars from the fibers. Carbohydr Polym 155:491–497. https://doi.org/10.1016/j.carbpol.2016.09.004

Park HA, Kubicki N, Gnyawali S, Chan YC, Roy S et al (2011) Natural vitamin E α-tocotrienol protects against ischemic stroke by induction of multidrug resistance-associated protein 1. Stroke 42(8):2308–2314

Parveez GKA, Christou P (1998) Biolistic-mediated DNA delivery and isolation of transgenic oil palm (*Elaeis guineensis* Jacq.) embryogenic callus cultures. J Oil Palm Res 10(2):29–38

Parveez GKA, Abrizah O, Nurhafizah R, Bahariah B (2010) Functional analysis of oil palm palmitoyl-ACP thioesterase (FatB) gene via down-regulation in a model plant: *Arabidopsis thaliana*. J Oil Palm Res 22:803–813

Parveez GKA, Rasid OA, Masani MYA, Sambanthamurthi R (2015) Biotechnology of oil palm: strategies towards manipulation of lipid content and composition. Plant Cell Rep 34(4):533–543

Perveen S (2018) Introductory chapter: terpenes and terpenoids. In: Perveen S, Al-Taweel A (eds) Terpenes and terpenoids. IntechOpen, London. https://doi.org/10.5772/intechopen.79683

Ping BTY, Gwendoline ECL (2006) Identification of lutein in crude palm oil and evaluation of carotenoids at various ripening stages of the oil palm fruit. J Oil Palm Res 18:189–197

Poblaciones MJ, Rodrigo SM, Santamaría O (2013) Evaluation of the potential of peas (*Pisum sativum* L.) to be used in selenium biofortification programs under mediterranean conditions. Biol Trace Elem Res 151(1):132–137

Pootakham W, Jomchai N, Ruang-Areerate P, Shearman JR, Sonthirod C et al (2015) Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). Genomics 105(5–6):288–295

Purba AR, Noyer JL, Baudouin L, Perrier X, Hamon S, Lagoda PJL (2000) A new aspect of genetic diversity of Indonesian oil palm (*Elaeis guineensis* Jacq.) revealed by isoenzyme and AFLP markers and its consequences for breeding. Theor Appl Genet 101(5–6):956–961

Radhika MS, Bhaskaram P, Balakrishna N, Ramalakshmi BA (2003) Red palm oil supplementation: a feasible diet-based approach to improve the vitamin A status of pregnant women and their infants. Food Nutr Bull 24(2):208–217

Rajanaidu N, Kushairi A, Rafii MY, Maizura I, Isa ZA et al (2006) Oil palm genetic resources and their utilization-a review. In: Proc. of Int. Symp. on Oil Palm Genetic Resources and their Utilization. Malaysian Palm Oil Board (MPOB), Kuala Lumpur, pp 33–80

Ramli US, Sambanthamurthi R (1996) β-ketoacyl synthase II in the oil palm (*Elaeis guineensis*) mesocarp. In: Williams JP, Khan MU, Lem NW (eds) Physiology, biochemistry and molecular biology of plant lipids, p 570

Ramzani PMA, Khalid M, Naveed M, Ahmad R, Shahid M (2016) Iron biofortification of wheat grains through integrated use of organic and chemical fertilizers in pH affected calcareous soil. Plant Physiol Biochem 104:284–293. https://doi.org/10.1016/j.plaphy.2016.04.053

Rance KA, Mayes S, Price Z, Jack PL, Corley RHV (2001) Quantitative trait loci for yield components in oil palm (*Elaeis guineensis* Jacq.). Theor Appl Genet 103(8):1302–1310

Rasid OA, Wan Nur Syuhada WS, Hanin AN, Masura SS, Zulqarnain M (2008) RT-PCR amplification and cloning of partial DNA sequence coding for oil palm (*Elaeis oleifera*) phytoene synthase gene. Asia Pac J Mol Biol Biotechnol 16(1):17–24

Rasid OA, Parveez GKA, Ho CY, Suhaimi N, Sambanthamurthi R (2009) Plant carotenoids: molecular genetics and regulation. J Oil Palm Res 21:588–601

Rasid OA, Syuhada W, Hanin A, Masura SS, Parveez GKA (2014) Molecular cloning and regulation of oil palm (*E. guineensis* Jacq.) phytoene desaturase in developing mesocarp tissues. J Oil Palm Res 26:37–46

Reddy MT, Kalpana M, Sivaraj N, Kamala V, Pandravada SR et al (2019) Indigenous traditional knowledge on health and equitable benefits of oil palm (*Elaeis* spp.). Open Access Libr J 6: e5103. https://doi.org/10.4236/oalib.1105103

Rodrigues-Neto JC, Correia MV, Souto AL, de Aquino Ribeiro JA, Vieira LR et al (2018) Metabolic fingerprinting analysis of oil palm reveals a set of differentially expressed metabolites in fatal yellowing symptomatic and non-symptomatic plants. Metabolomics 14:142. https://doi.org/10.1007/s11306-018-1436-7

Rodríguez-García C, Sánchez-Quesada C, Toledo E, Delgado-Rodríguez M, Gaforio JJ (2019) Naturally lignan-rich foods: a dietary tool for health promotion? Molecules 24(5):917. https://doi.org/10.3390/molecules24050917

Römer S, Fraser PD, Kiano JW, Shipton CA, Misawa N, Schuch W, Bramley PM (2000) Elevation of the provitamin A content of transgenic tomato plants. Nat Biotechnol 18(6):666–669

Rosli R, Ab Halim MA, Chan KL, Azizi N, Jayanthi N (2014) Genomsawit website, MPOB information series no 134. Malaysian Palm Oil Board, Kuala Lumpur

Rosli R, Amiruddin N, Ab Halim MA, Chan PL, Chan KL et al (2018a) Comparative genomic and transcriptomic analysis of selected fatty acid biosynthesis genes and CNL disease resistance genes in oil palm. PLoS One 13(4):e0194792. https://doi.org/10.1371/journal.pone.0194792

Rosli R, Chan PL, Chan KL, Amiruddin N, Low ETL et al (2018b) In silico characterization and expression profiling of the diacylglycerol acyltransferase gene family (DGAT1, DGAT2, DGAT3 and WS/DGAT) from oil palm, *Elaeis guineensis*. Plant Sci 275:84–96

Rosli R, Ab Halim MA, Ting NC, Mohd Zaki N, Jayanthi N et al (2022) Oil palm SSR resource interface (OPSRI) –web-based bioinformatic analysis pipeline for SSR mining. J Oil Palm Res. https://doi.org/10.21894/jopr.2022.0008

Rozali NL, Tahir NI, Hassan H, Othman A, Ramli US (2021) Identification of amines, amino and organic acids in oil palm (Elaeis guineensis Jacq.) spear leaf. Biocatal Agric Biotechnol 37 (102165). https://doi.org/10.1016/j.bcab.2021.102165

Samant GV, Wali VB, Sylvester PW (2010) Anti-proliferative effects of gamma-tocotrienol on mammary tumour cells are associated with suppression of cell cycle progression. Cell Prolif 43: 77–83

Sambanthamurthi R, Abdullah SNA, Kadir APG (2002) Genetic manipulation of the oil palm-challenges and prospects. Planter 78:547–564

Sambanthamurthi R, Tan Y, Sundram K, Abeywardena M, Sambandan TG et al (2011) Oil palm vegetation liquor: a new source of phenolic bioactives. Br J Nutr 106(11):1655–1663. https://doi.org/10.1017/S0007114511002121

Sanusi NSNM, Rosli R, Ab Halim MA, Chan KL, Jayanthi N (2018) PalmXplore: oil palm gene database. Database 2018:bay095. https://doi.org/10.1093/database/bay095

Sathya A, Vijayabharati R, Srinivas V, Gopalakrishnan S (2013) Plant growth- promoting action-bacteria on chickpea seed mineral density: an upcoming complementary tool for sustainable biofortification strategy. 3Biotechnol 6(2):138. https://doi.org/10.1007/s13205-016-0458-y

Sen CK, Khanna S, Roy S (2007) Tocotrienols in health and disease: the other half of the natural vitamin E family. Mol Asp Med 28(5–6):692–728

Shah FH, Rashid O, Simons AJ, Dunsdon A (1994) The utility of RAPD markers for the determination of genetic variation in oil palm (*Elaeis guineensis*). Theor Appl Genet 89(6): 713–718

Shah FH, Rashid O, San CT (2000) Temporal regulation of two isoforms of cDNA clones encoding delta 9-stearoyl–ACP desaturase from oil palm (*Elaeis guineensis*). Plant Sci 152(1):27–33

Shah S, Gapor A, Sylvester PW (2003) Role of caspase-8 activation in mediating vitamin E-induced apoptosis in murine mammary cancer cells. Nutr Cancer 45(2):236–246

Shilpa K, Varun K, Lakshmi BS (2010) An alternate method of natural drug production: eliciting secondary metabolite production using plant cell culture. J Plant Sci 5:222–247. https://doi.org/10.3923/jps.2010.222.247

Shin MG, Ithnin M, Vu WT, Kamaruddin K, Chin TN et al (2021) Association mapping analysis of oil palm interspecific hybrid populations and predicting phenotypic values via machine learning algorithms. Plant Breed 140(6):1150–1165

Singh R, Tan SG, Panandam JM, Rahman RA, Ooi LCL et al (2009) Mapping quantitative trait loci (QTLs) for fatty acid composition in an interspecific cross of oil palm. BMC Plant Biol 9:114. https://doi.org/10.1186/1471-2229-9-114

Singh R, Ong-Abdullah M, Low ETL, Abdul Manaf MA, Rosli R et al (2013a) Oil palm genome sequences reveal divergence of interfertile species in old and new worlds. Nature 500(7462): 335–339

Singh R, Low ETL, Ooi LC-L, Ong-Abdullah M, Ting NC (2013b) The oil palm *SHELL* gene controls oil yield and encodes a homologue of SEEDSTICK. Nature 500:340–344

Singh R, Low ETL, Ooi LCL, Ong-Abdullah M, Ting NC et al (2014) The oil palm VIRESCENS gene controls fruit colour and encodes a R2R3-MYB. Nat Commun 5:4106

Siti Nor Akmar A, Cheah SC, Nurniwalis AW (2014) Expression regulatory elements. United States Patent No US 8,791,330 B2

Sivan YS, Jayakumar YA, Arumughan C, Sundaresan A, Jayalekshmy A et al (2002) Impact of vitamin A supplementation through different dosages of red palm oil and retinol palmitate on preschool children. J Trop Pediatr 48(1):24–28

Sowmya C, Venu MVB, Kalyana B (2017) Identification of polymorphic microsatellite markers for (*Elaeis guineensis* Jacq.). Int J Oil Palm 9(1&2):17–20

Sundram K, Khor HT, Ong AS, Pathmanathan R (1989) Effect of dietary palm oils on mammary carcinogenesis in female rats induced by 7,12-dimethylbenz(a)anthracene. Cancer Res 49: 1447–1451

Syahanim S, Abrizah O, Siti Nor Akmar A, Mohamad Arif AM, Ho CL (2007) Cloning of an oleoyl-Coa desaturase from oil palm. In: Proc. of the PIPOC 2007 Int. Palm Oil Congress (Agriculture, Biotechnology & Sustainability). Malaysian Palm Oil Board, Kuala Lumpur, pp 1001–1009

Syarifah-Noratiqah SB, Zulfarina MS, Ahmad SU, Fairus S, Naina-Mohamed I (2019) The pharmacological potential of oil palm phenolics (OPP) individual components. Int J Med Sci 16(5):711–719. https://doi.org/10.7150/ijms.29934

Sylvester PW, Shah SJ (2005) Mechanisms mediating the antiproliferative and apoptotic effects of vitamin E in mammary cancer cells. Front Biosci-Landmark 10(1):699–709

Tahir NI, Shaari K, Abas F, Parveez GKA, Zamzuri I, Ramli US (2012) Characterization of apigenin and luteolin derivatives from oil palm (Elaeis guineensis Jacq.) leaf using LC-ESI-MS/MS. J Agric Food Chem 60(45):11201–11210. https://doi.org/10.1021/jf303267e

Tahir NI, Shaari K, Abas F, Parveez GKA, Hashim AT (2013) Identification of oil palm (*Elaeis guineensis*) spear leaf metabolites using mass spectrometry and neutral loss analysis. J Oil Palm Res 25(1):72–83

Tahir NI, Rozali NL, Rahmah ARS, Amiruddin MD, Lim FH et al (2022) Metabolome study of oil palm (*Elaeis guineensis* Jacq.) planted in different environment conditions. Trop Plant Biol 15(3). https://doi.org/10.1007/s12042-022-09318-6

Tan CH, Ghazali HM, Kuntom A, Tan CP, Ariffin AA (2009) Extraction and physicochemical properties of low free fatty acid crude palm oil. Food Chem 113(2):645–650

Tardugno R, Cicero N, Costa R, Nava V, Vadalà R (2022) Exploring lignans, a class of health promoting compounds, in a variety of edible oils from Brazil. Foods 11:1386. https://doi.org/10.3390/foods11101386

Teh CK, Ong AL, Kwong QB, Apparow S, Chew FT et al (2016) Genome-wide association study identifies three key loci for high mesocarp oil content in perennial crop oil palm. Sci Rep 6(1): 1–7

Ting NC, Zaki NM, Rosli R, Low ETL, Ithnin M et al (2010) SSR mining in oil palm EST database: application in oil palm germplasm diversity studies. J Genet 89(2):135–145

Ting NC, Jansen J, Mayes S, Massawe F, Sambanthamurthi R et al (2014) High density SNP and SSR-based genetic maps of two independent oil palm hybrids. BMC Genomics 15:309

Ting NC, Yaakub Z, Kamaruddin K, Mayes S, Massawe F et al (2016) Fine-mapping and cross-validation of QTLs linked to fatty acid composition in multiple independent interspecific crosses of oil palm. BMC Genomics 17:289

Ting NC, Mayes S, Massawe F, Sambanthamurthi R, Chan KL et al (2021) Candidate genes linked to QTL regions associated with fatty acid composition in oil palm. Biologia 76:267–279. https://doi.org/10.2478/s11756-020-00563-2

Tow WK, Goh APT, Sundralingam U, Palanisamy UD, Sivasothy Y (2021) Flavonoid composition and pharmacological properties of *Elaeis guineensis* Jacq. leaf extracts: a systematic review. Pharmaceuticals 14(10):961

Tsouko E, Alexandri M, Fernandes KV, Freire DMG, Mallouchos A, Koutinas AA (2019) Extraction of phenolic compounds from palm oil processing residues and their application as antioxidants. Food Technology and Biotechnology 57(1):29–38. https://doi.org/10.17113/ftb.57.01.19.5784

Turner PD (1981) Oil palm diseases and disorders. Oxford University Press

Van Stuijvenberg E, Faber M, Dhansay MA, Lombard CJ, Vorster N, Benadé AJSM (2000) Red palm oil as a source of β-carotene in a school biscuit used to address vitamin A deficiency in primary school children. Int J Food Sci Nutr 51(1):S43–S50

Venu MVB, Prasanna Kumar G, Kalyana B (2018) Molecular characterization of oil palm (*Elaeis guineensis* Jacq.) germplasm using microsatellite markers. Int J Oil Palm 10(1&2):13–16

Wada S (2009) Chemoprevention of tocotrienols: the mechanism of antiproliferative effects. Forum Nutr 61:204–216

Winkel BSJ (2004) Metabolic channeling in plants. Annu Rev Plant Biol 55:85–107. https://doi.org/10.1146/annurev.arplant.55.031903.141714

Woittiez LS, Van Wijk MT, Slingerland M, Van Noordwijk M, Giller KE (2017) Yield gaps in oil palm: a quantitative review of contributing factors. Eur J Agron 83:57–77

Wu G (2013) Functional amino acids in nutrition and health. Amino Acids 45:407–411. https://doi.org/10.1007/s00726-013-1500-6

Xia W, Luo T, Dou Y, Zhang W, Mason AS et al (2019) Identification and validation of candidate genes involved in fatty acid content in oil palm by Genome-Wide Association analysis. Front Plant Sci 10:1263. https://doi.org/10.3389/fpls.2019.01263

Yaakub Z, Singh R, Ting NC, Marhalil M, Xaviar A et al (2020) Quantitative trait loci (QTL) linked to compactness in an interspecific backcross two (BC2) population of oil palm. J Oil Palm Res 32(2):201–210

Yamanushi TT, Ichinose T, Seto H, Sagai M, Torii MI, Igarashi O (2001) The effect of dietary carotenoids on lung tumorigenesis induced by intratracheally instillated diesel exhaust particles. J Nutr Sci Vitaminol 47:32–39

Yeap WC, Khan NCM, Jamalludin N, Muad MR, Appleton DR et al (2021) An efficient clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein 9 mutagenesis system for oil palm (*Elaeis guineensis*). Front Plant Sci 12:773656. https://doi.org/10.3389/fpls.2021.773656

Yin NS, Abdullah ST, Phin CK (2013) Phytochemical constituents from leaves of *Elaeis guineensis* and their antioxidant and antimicrobial activities. Int J Pharm Pharm Sci 5:137–140

Yuan L, Wu L, Yang C, Quin LV (2013) Effects of iron and zinc foliar applications on rice plants and their grain accumulation and grain nutritional quality. J Sci Food Agric 93(2):254–261. https://doi.org/10.1002/jsfa.5749

Zagré NM, Delpeuch F, Traissac P, Delisle H (2003) Red palm oil as a source of vitamin A for mothers and children: impact of a pilot project in Burkina Faso. Public Health Nutr 6(8):733–742

Zhang YQ, Pang LL, Yan P, Liu DY, Zhang W et al (2013) Zinc fertilizer placement affects zinc content in maize plant. Plant Soil 372:81–92. https://doi.org/10.1007/s11104-013-1904-9

Zubaidah R, Abdullah SNA (2010) Functional characterisation of the oil palm type 3 metallothionein-like gene (MT3-B) promoter. Plant Mol Biol Report 28(3):531–541

Zubaidah R, Nurniwalis AW, Siti Nor Akmar A, Ghulam Kadir AP (2017) The use of *Arabidopsis thaliana* model system for testing oil palm promoter: case study on oil palm MT3-A promoter. J Oil Palm Res 29(2):189–196

Part III

Pulse Crops

Nutritional Traits of Beans (*Phaseolus vulgaris*): Nutraceutical Characterization and Genomics

M. W. Blair, H. Li, L. Nekkalapudi, V. Becerra, and M. Paredes

Contents

Abstract

Common bean is among the top three edible legumes on the planet and is very diverse in seed size and shape constituting two genepools (Andean and Mesoamerican) and many commercial seed types within each of these. The seed structure of common bean provides a basis for defining some of the nutraceutical

Note: Authors V. Becerra and M. Paredes are retired from INIA - Instituto de Investigaciones Agropecuarias

M. W. Blair (✉) · H. Li · L. Nekkalapudi
TSU – Tennessee State University, Department of Agricultural and Environmental Sciences, Nashville, TN, USA
e-mail: mblair@tnstate.edu

V. Becerra
INIA – Instituto de Investigaciones Agropecuarias, Chillán, Chile

M. Paredes
INIA – Instituto de Investigaciones Agropecuarias, Chillán, Chile

Univ Vina del Mar, Escuela de Ciencas Agricolas y veterinarias, Vina del Mar, Chile

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_23

properties of the crop: The seed coat while less than a tenth of the seed weight is packed with secondary metabolites that can have health benefits ranging from anti-cancer effects to diabetes control. Since the seed coat is maternal tissue, it is important that fixed lines be considered so that the hybrid internal tissues match the outside genotype. The cotyledonary tissue is a powerhouse of macronutrients, predominantly storage proteins, fiber, and complex starches that can control obesity and contribute to reduction in heart disease. Minerals and vitamins make up micronutrients that are found in the embryonic axis and other tissues of common beans providing health benefits especially to women and infants having iron deficiency anemia and other hidden hunger effects. Regular consumption of beans can have multiple nutritional benefits, for example, phytosterols can reduce blood cholesterol levels and hypertension. High calcium and magnesium are important for bone strength and cardiovascular function. The genetics behind nutraceutical properties of common bean are not well studied and require a detailed analysis of seed physiology, growth, and development and genotypic differences for each trait described above. Transcriptomics can be as useful as genomics in this analysis. Therefore, we propose a tissue-specific, gene expression-based analysis for improvement of certain nutraceutical properties using as examples the well-known flavonoid pathway that leads to many beneficial molecules found in the different tissues of common beans, as well as starch, phytate, and protein pathways. Since this pulse is mostly consumed directly after boiling, common beans do not require the extensive processing that some other grain legumes or pulses undergo, making their analysis for health benefits easier to evaluate. However, given many seed types, colors, and origins, researchers should use multiple carefully selected genotypes to make conclusions about specific commercial classes of beans or nutritional traits.

Keywords

Commercial bean classes · Cotyledon storage proteins · Dry grain pulse · Embryo axis · *Phaseolus* genus · Seed coat colors

1 Introduction

Nutraceuticals are those plant compounds that have a beneficial effect on human health. They contribute to overall food security but also determine the food quality of plant products such as grains. They include macronutrients, micronutrients, and secondary metabolites unique to each group of plants. High-quality, nutritious pulses (grain legumes) are one key to food security in the world's human population (Carbonaro et al. 2015). The legumes are on par in importance with calorie-supplying cereals but are the major source of protein among income-limited groups and vegetarian/vegan consumers. Pulses are major suppliers of higher protein levels in grain-based diets along with essential amino acids such as lysine and tryptophan that are low in cassava, maize, potatoes, or rice, the main crops for carbohydrate supply in the world.

Among legumes, the common bean (*Phaseolus vulgaris* L.) is considered the most important for direct human consumption in terms of total area of production, amounts consumed, and contribution to nutritional quality of various diets across many nations (Hayat et al. 2014) with various reviews written on their potential for biofortification and essential nutritional compounds (Blair 2013; Nyau 2014). Given their diversity, it is important to know that common beans were domesticated in two sections of the New World, Central/North America and South America, and thus have two major genepools. Common beans are therefore a very diverse crop with many different seed types depending on the environmental zone in which the crop is grown and the consumers supplied with beans of various commercial classes.

Small and medium seeded beans from Central/North America are termed Mesoamerican, while those from South America are more specifically called Andean based on their origin along the Andes Mountains. Among the Andean genepool are the commercial classes of kidney beans (pink and red), popping beans (nuna), red mottled (Cargamanto rojo, Pompadour, and Rosecoco), sugar beans (Borlotto, Cargamanto blanco, Coco, and cranberry types), and yellow beans (Canario and Jalo) plus others used on more local levels, such as Liborino a yellow Cargamanto bean (Peláez et al. 2022) found only in one region of Colombia. Andean beans became prevalent in Eastern and Southern Africa to which they are not native but where production and consumption levels are very high.

Among the Mesoamerican genepool are the commercial classes of black beans, navy beans, pintos, small reds, all major classes across various countries, and some yellows or tans. Carioca and Jalinho are also small seeded and Mesoamerican beans found originally in Brazil, with some expansion into Africa. Various specialty types are found only in Guatemala and Mexico such as shiny black seeded or Flor de Mayo/Junio types, respectively. Great Northern beans are medium-sized white beans from the Plains region of the United States. Figure 1 shows a range of these bean types with variation of seed color and size.

As a major export crop, common beans are traded globally and regionally from country to country and zone to zone. At the national level, local trading from rural to urban areas is very important and ensures fresh harvested beans. Exported beans tend to be less fresh and require months in transit from a producer region often in the northern hemisphere (Canada, China, and the United States) to the tropics and sub-tropics where consumption is higher than in countries of origin. For example, US beans arrive in Central America from ports in the upper Mississippi basin, while Canadian beans reach Cuba from the Atlantic Ocean. Chinese beans are shipped worldwide, and some arrive in Latin America or the Middle East. North American beans are prevalent in European markets. Exports/imports between neighboring countries of Argentina/Bolivia and Brazil are also important in terms of trade volume. Total worldwide production is between 25 and 30 million (M) metric tons grown on up to 35M hectares according to data available for the past decade (www.fao.org/faostat). A greater amount of Mesoamerican genepool beans than Andean genepool beans are produced when considering all production regions but the latter tend to be produced more locally and are less a part of the export, although exceptions to this are found for certain growing regions and consumer markets.

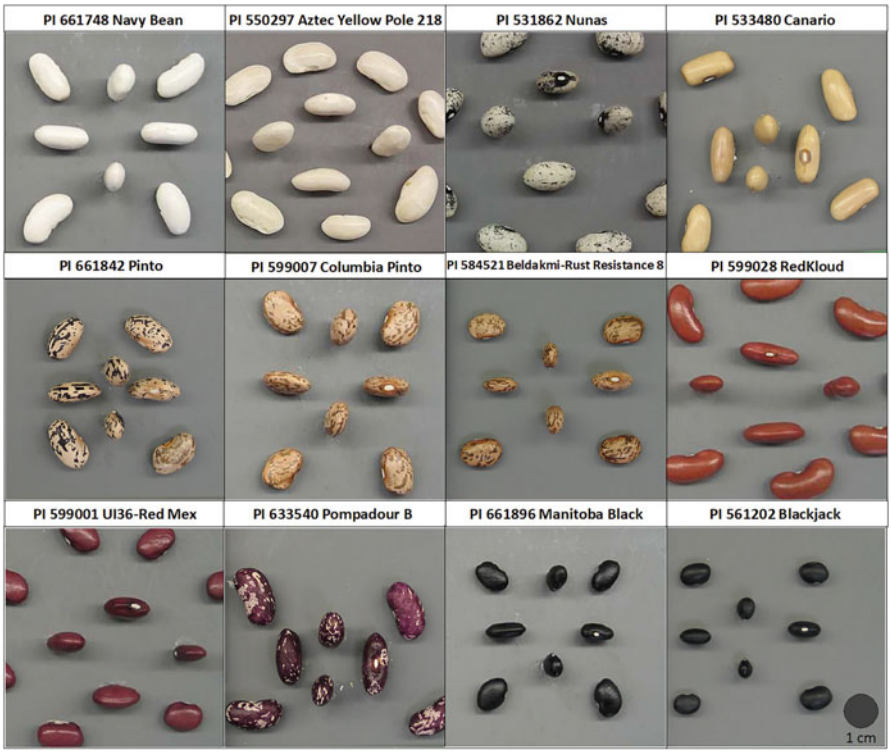


Fig. 1 Pictures of the colors of common beans based on plant introductions (PIs) shown to scale from the Genetic Resources Information Network (GRIN) of the USDA

Some inaccuracy in data for Asia may reflect confusion between *Phaseolus* beans and some *Vigna* species, although we only discuss the former genus. This chapter covers nutraceuticals in the common bean per se with highlighting overall characteristics of their contribution to diets.

2 Races/Subraces of Common Bean

Commercial classes of common bean align with genotypically and phenotypically defined races and subraces of the species (Blair et al. 2006). For example, race Durango-Jalisco, joined as races by some authors but somewhat distinct in plant morphology, include almost all the medium-sized beans originating mostly in Mexico. Race Mesoamerica includes most small seeded beans of different colors (Díaz and Blair 2006). Subraces have been proposed for this race with small red beans proposed as subrace Honduras, black beans grouping as subrace Veracruz, and various light-colored beans as subraces Carioca and Jalinho (Blair et al. 2013a, b).

Andean beans are divided into two major races (Nueva Granada and Peru) with a third subgroup (Chile) and have large diversity in seed colors making it difficult to distinguish commercial classes and with substantial overlaps in races as well. The importance of commercial classes and races/subraces of common bean lies in their association with some nutraceutical and plant morphology characteristics as well as each having a limited number of alleles for genes of interest. Common bean is among the most diverse in terms of number of seed types with colors ranging from white to yellow, tan, brown, pink, red, purple, and black with varying sizes. Seed size varies with the genepools, and races as discussed above determine the variability in many plant organs within the seed or those organs derived form it in seedling stages. For example, the sizes of cotyledon, embryo axis, and cotyledonary/primary leaves are all correlated with seed size and vice versa are influenced by these factors.

Proportion of seed coat as maternal tissue to non-seed coat as filial tissue is also related to seed size as we will discuss below. These relationships have implication on the nutraceutical properties of common bean genotypes, as different chemistries are stored in each part of the grain that we consume and even though the grain is consumed whole for the most part, the proportion of each organ influences where primary and secondary metabolites or vitamins and minerals accumulate.

3 Seed Organs and Nutritional Quality

The seed of grain legumes like common beans, being dicotyledonous, are made up of embryo and two cotyledons which are enlarged parallel storage organs that replace leaves from the growing hybrid plant as nutrient-rich tissues (Gomes et al. 2018). Figure 2 summarizes the relationship of the seed organs with nutraceutical properties as discussed in the next section using red to indicate anti-nutritional properties and green beneficial properties. The embryo is a small proportion of the total seed weight occupying a space in between the large cotyledons that make up the majority of the seed concentration. The embryo is a powerhouse of nutraceutical vitamins, hormones, and minerals with benefits listed in the niche squares above and to the upper left of the bean in the figure below. Cotyledons have both anti-nutritional (lower left, red) and beneficial (lower right, green) nutraceutical characteristics.

The seed coat is a maternal tissue derived from the mother plant that produced the seed. In an inbreeding crop like beans, this differentiation is only important if two genotypes are hybridized or if a plant happens to outcross, but it does indicate distinctions of the pulses compared to the grains of monocotyledonous cereals and have significant implications on the accumulation of nutrients, secondary metabolites, and nutraceuticals across the seed itself. Below we will discuss various macronutrients, micronutrients, and health-promoting factors in relationship to where they occur in the seed, so we refer the reader to the next figures summarizing the health effects found in each part of the seed (Figs. 3 and 4). Since common bean is generally consumed whole and not decorticated, the different locations of nutrients and nutraceuticals have implications on plant physiology, gene expression within the

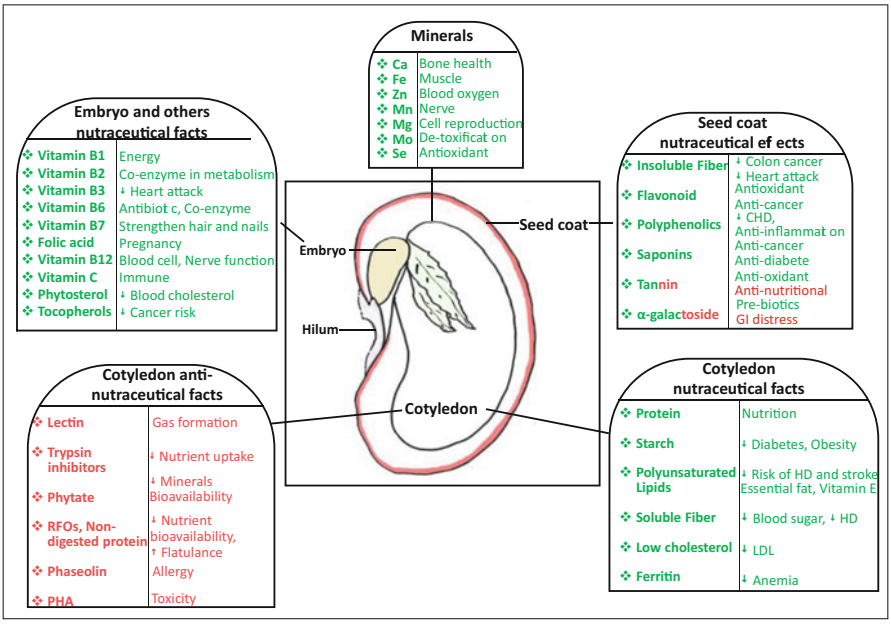


Fig. 2 Distribution of nutraceutical properties among common bean seed organs

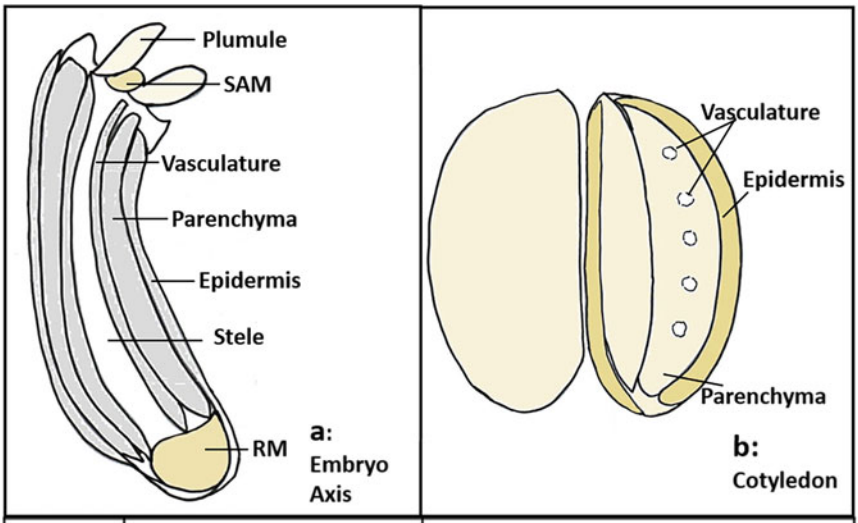
seed, and bioavailability to the human diet, more than to methods of preparation or potential processing of this pulse crop.

4 Macronutrients: Proteins, Carbohydrates, and Lipids

As pulse legumes, common beans are complementary to non-legume grains because they provide high amounts of proteins and specific amino acids that cereals are low in. According to OECD (2019), common beans of different commercial classes range from around 23% protein (in Navy and pink Mesoamericans) to over 26% (in Cranberry and kidney Andean types).

Specific authors have estimated a range in protein concentration from as low as 18% to as high as 30% for common beans (Rodiño et al. 2003). While this is often related to the variety involved, genotype x environmental effects are significant (Florez et al. 2009). Low yield environments can produce common beans of higher protein concentration, but so can higher yield environments; and nitrogen fixation may be one component of protein accumulation in common bean as a pulse grain. The types of seed storage proteins and their relative abundance can also influence overall protein quantity and quality in terms of digestibility (Osborn et al. 2003; Emani and Hall 2008).

Phaseolin, a type of globulin, is usually most of the protein in the cotyledon (40 to 60%), but the embryo axis can be filled with many other types of polypeptides (Carbonaro 2006). Albumins, with 10–20%, are the next most common seed protein. Prolamins are much smaller portions of total proteins, representing 2% or less.



Proteins	✓ High quality diverse proteins ✓ Balanced amino acid profile ✓ High sulfur containing amino acid ✓ Ferritin/ Fe in axis	✓ Storage proteins predominant ✓ Globulins / Phaseolin ✓ Ferritin/ Fe in vasculature
Carb	✓ Low Fiber ✓ Digestible sugars	✓ High Fiber ✓ Resistant starches ✓ Sugar (Glucose, raffinose, stachyos)
Phytate	Phytate lower	Phytate concentration high
Minerals	Zinc containing enzymes and Potassium	Zinc and Iron bioavailability are reduced by phytate

Fig. 3 Structures of (**a**) embryo axis and (**b**) cotyledons as seed tissues and their nutrient traits

However, both these proteins are important sources of essential sulfur-containing amino acids for common beans, which are considered low in these. Related to mineral content, ferritin is an iron-storing protein that is found mainly in the vascular tissues of cotyledons and the embryo axis within the seed (Cvitanich et al. 2010). It is determinative in how much iron is in the seed and where it is located. Anti-digestive proteins include lectins, phytohemagglutinins, and protease inhibitors, but these are often broken down by cooking of common beans (De Mejía et al. 2003).

Common beans are also high in carbohydrate concentration as this is the principal nutritional component of this type of pulse being up to 70% of the dry grain (OECD 2019). Bean carbohydrates can be divided into sugars and starches (Rivera et al. 2018). The simple sugars are glucose that is easily digestible, while

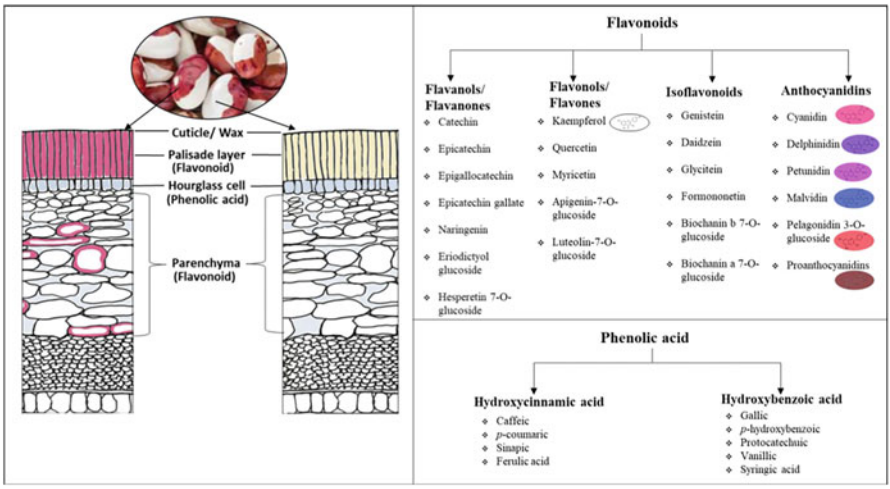


Fig. 4 Diagram of seed coat structure and the color containing portions (to the left) and uncolored portions (to the right) of this seed tissue. Seed coat flavonoids and phenolic acid component in common beans

complex sugars such as raffinose, stachyose, and verbascose are not (Sathe et al. 1984). They are considered flatulence causing factors in common bean as they are not digested in the stomach but in the lower intestine where they cause discomfort! The starches in common bean include amylose and amylopectin (Rivera et al. 2018) which are found in the two cotyledons usually in starch crystals inside specialized accumulator cells (Uebersax et al. 2022). Common beans are high in fiber, which is found mainly in the seed coat (Dhital et al. 2016; Ma et al. 2017). Fiber in beans is mostly not digestible because it is locked inside the hardened cell walls of seed coat tissue cells.

In contrast to proteins and carbohydrates, common beans are very low in lipid content. They are never pressed for oil, unlike peanuts or soybeans, two other pulses that are used for extracting oils. The small amount of fats is of high quality though being mostly polyunsaturated. Overall concentration of lipids is 1–2% among major commercial classes tested by OECD (2019) or in Spain (Rivera et al. 2018). Common bean lipids are from cell membranes and vesicles of cotyledon and embryo tissues. Like proteins, lipids are not common in the seed coat, but waxes occur on that tissue and determine if the grain type is shiny (low wax) or matte (high wax).

5 Micronutrients: Minerals and Vitamins

Common bean is a superstar crop in terms of certain essential minerals and vitamins. Since these are required by human diets but in small (micro) amounts, they are often called essential micronutrients. Among the minerals are Fe, iron; S, sulfur; Se,

selenium; and Zn, zinc all targets of biofortification or breeding for enhanced concentration above the normal or average amount found in common bean seed (Blair 2013). Soil effects are significant on the accumulation of Zn and Se, while homeostasis is active to prevent over-accumulation of Fe beyond what is needed for photosynthesis and other enzymatic processes. The amount of S usually depends on sulfur-containing amino acids and the proportion of different proteins, while the amount of Se is highly dependent on the soils on which the beans grow.

Meanwhile Fe and Zn have been studied well for inheritance and are genotype-dependent with factors of partitioning and environmental interactions coming into play to determine whether a variety is biofortified or not. Both elements can be enhanced by micro-fertilization, especially Zn which tends to be deficient in soils used to grow common beans but can easily be absorbed from foliar or seed treatments. In addition, common beans contain small amounts of other microelements such as B, boron, and Cu, copper.

The macroelements, N, nitrogen, and P, phosphorus, are related to limited supply in unfertilized soil or the amount provided as fertilizer. While N is proportional to the protein content of seed, P is often closely related to an antinutrient to the human diet which is phytate. The availability of these two elements for common bean depends on soil fertility conditions such as organic matter concentration as well as on biological/symbiotic nitrogen fixation (SNF) which in turn is highly dependent on the supply of organic/inorganic P cycles in the soil. The amount of Ca, calcium; K, potassium; Mg, magnesium; and Mn, manganese are related to soil conditions and plant uptake plus homeostasis, as is the case for Fe as well but less so for Zn. Genomics regions based on quantitative trait locus (QTL) studies for all these minerals have been detected providing a selection method to reach desired value of mineral concentration, except where correlations inhibit selection for opposing factors (Blair et al. 2016; Izquierdo et al. 2018).

The division of micro- and micronutrient between different seed organs varies based on the mineral or element. For example, Ca along with Mg are most abundant in the seed coat, with up to 85% found there. Phytates are abundant in the cotyledons (Moraghan and Grafton 2002) and a portion of the Fe of the seed, especially that portion found as ferritin. Zn and K are found principally in the embryo axis, as is another portion of Fe (Moraghan and Grafton 2002). In turn, the amount of S is related to the prevalence of phaseolin and lectins which are low in this microelement versus other proteins generally found in the embryo axis which are high in it. Many of the microelements are found as part of the embryo where there is a greater diversity of enzymes that use them as co-factors.

Common bean grains are rich sources of certain vitamins mainly those in the vitamin B family including folates (a.k.a vitamin B9), niacins (vitamin B3), thiamine (vitamin B1), riboflavins (vitamin B2), pantothenic acid (vitamin B5), and vitamin B6 (OECD 2019). Niacins are of the highest concentration in parts per million (ppm) of dry seed (mg/Kg) being 5.4–26.8 ppm. Thiamine are next being 4.9–11.4 ppm. Pantothenic acid follows with 3.5–12.3 ppm. Others are vitamin B6 (3.2–5.7 ppm), vitamin B9 (1.2–5.8 ppm), and vitamin B2 (1.4–2.9 ppm). In contrast, common bean grains are low in vitamin A or C. These vitamins are only found in leaves of the crop

that are sometimes cooked as a vegetable dish, primarily in sub-Saharan Africa. The low lipid content of the pulse means that the latter vitamins are not highly available, so it is common for beans to be eaten with other sources of vitamins such as pumpkin, carrots, and tomatoes which increase absorption of certain minerals such as Fe and Zn.

6 Other Health-Promoting Factors

The health benefits of a diet rich in pulses have long been known to scientists and some consumers but are better studied for European species such as chickpeas compared to those grain legumes popular in Africa or the Americas such as common bean (Nyau 2014). Various factors of bean seeds are health promoting, but we will limit this review to those that have been studied for common bean.

The health benefits of common beans address some of the major diseases of developed and developing countries including diabetes, certain cancers, iron deficiency anemia, micronutrient deficiencies, and heart disease/stroke risk (Anderson and Major 2002). Obesity is of growing concern around the world and can be partially controlled by pulse-rich diets as these give a feeling of satiety that reduces overall calorie intake in some cases. Interactions of various conditions means that lower blood pressure and lower obesity translate into lower risk of cardiac disease, heart attacks, and strokes.

Diabetics can benefit from eating common beans mainly because they are low in available sugars and therefore are of low glycemic index. The low glycemic index associated with beans may also be due to their complex starches or high amounts of fiber ranging from 8% to 28% of seed (Singh et al. 2017). Complex starches have been discussed and refer to non-storage carbohydrates in seeds. They include disaccharides and oligosaccharides that are difficult to digest in the stomach and intestine but which are digested by bacteria in the lower gut leading to flatulence. They play a role in gut health as they reach the colon intact and encourage beneficial bacteria to grow.

Fiber contents often depend on commercial class (OECD 2019) with this fraction consisting of cellulose, hemicellulose, mucilage, and pectins along with digestion-resistant starch making beans healthy and useful to control diabetes (Mojica et al. 2017). Insoluble fibers are generally found in the seed coat, while fiber from the cotyledons have various characteristics of water solubility (Wang and Toews 2011). A smaller component of fiber-derived products is digestion-resistant starches or complex sugar such as beta-glucans and inulin (Rebello et al. 2014). Inulin has been postulated to be a promoter of gut health and micronutrient uptake. Fiber amount in the seed coat depends on the seed coat thickness which appears to vary in different genepools and different varieties. This fiber fraction can correlate with longer cooking time making it a balancing act among consumer preferences for fast cooking and health benefits of higher fiber. However, consumption is encouraged for diabetics and those at heart attack risk, despite the longer cooking time that can modify other components of the final cooked bean product.

Another health benefit of beans is their postulated reduction in certain cancer prevalence. Most studies have concentrated on breast and colon cancers and are in mouse and rat models for these diseases rather than in human clinical studies (Reynoso Camacho et al. 2007). The beneficial effect of common bean on human health depends on their rate of consumption, with healthy diets usually including 20 Kg or more of pulses per year. High consumption is also needed for common bean biofortification of diets to be effective (Blair 2013). However, the absorption of Fe and Zn can be influenced by phytate concentration so that in addition to being a macronutrient, amount of P is surrogate for micronutrient availability.

Anti-cancer affects have also been ascribed to the flavonoids and other phenolics such as tannins in common beans. Suggestive of this is the lower incidence of colon cancer in Guatemala where black beans are commonly consumed. However, low tannin/low flavonoid beans that are white in seed coat color are often among the best beans for in vitro anti-cancer assays for other cancer types such as breast and prostate cancers. Therefore, the food components and exact metabolic compounds behind the anti-cancer effects of common beans are still a mystery. In this review, we look at the components in the seed coat and which are associated with seed color as shown in Fig. 4. This may explain the diversity of seed types selected by consumers over many locations and many generations.

Quercetin, a component of tannins and part of the flavonoid synthesis pathway (to be discussed later), has been suggested to have anti-inflammatory characteristics that blocks cyclooxygenase pathway and protects against heart disease and attacks (Nijveldt et al. 2001). Other secondary metabolites from common bean seed coats such as pro-anthocyanins can act as antioxidants reducing effects of detrimental cholesterol. Again, these are concentrated in the seed coat, specifically in palisade layer and certain parenchyma cells that show color. A layer of hourglass cells below the palisade layer has no color components but can accumulate tannins that lead to darkening of the seed coat with time and exposure to UV light. This darkening is well understood to be controlled by a few major genes and to involve enzymatic reactions that combine phenolics into complex tannins.

Common beans also contain phytosterols that are similar in structure to cholesterol but as they are plant derived have health benefits of reducing serum LDL levels and help manage blood cholesterol levels in human beings. They do this in the intestine where the phytosterols mix with bile salts displacing animal-derived cholesterols, inhibiting their absorption or retention by the intestinal lining and allowing these to be excreted into the eventual waste stream. Tannins, phytosterols, and flavonoids may act in conjunction with the high fiber of whole grains such as common bean as well as associated factors from a healthy diet to reduce heart attack risk (Jeon et al. 2012). A variety of bean types are eaten in Mediterranean diets that correlate with slow aging foods, low rates of diabetes, cancers, and obesity. Other bioactive carbohydrates apart from fiber and starch may also play a role in reducing diabetes, coronary disease, or cancer risk either directly or through the gut bioflora that common bean consumption encourages with non-starch poly-, oligo-, and monosaccharides. Slow digestion of some of the polysaccharides found in common bean can control blood glucose levels compared to rapidly digested starches found in

most cereal products. Resistant starches and fiber pass through the large intestine where they can control colon pH levels, increase fecal matter, and affect any further digestion processes along the way.

Slow digest starches can account for 25% of all polysaccharides and are principally found in the cotyledonary portion of the seed (Yadav et al. 2010; Ma et al. 2017). Another mechanism for heart health can be the lowering of cholesterol and other heart damaging lipids. Overall the benefits of common bean start with reduction in the onset of obesity which helps with heart health and minimizing type II diabetes. In countries where common bean consumption was traditional but is declining, such as Puerto Rico, Mexico, Guatemala, and Colombia, we see higher rates of this diabetes and people being overweight. Consumption rates in the United States are too low to see beneficial effects of bean-based diets except with certain communities that are vegan or vegetarian, such as Adventists, who have longer life spans than others. Previous research found that people who consume legumes at least four times per week have 22% reduced risk of heart diseases.

7 Antinutrients: Lectins, Phytates, Phytohemagglutinins, Tannins, Trypsin Inhibitors

A fraction of protein in common bean is associated with anti-herbivory and as such is anti-nutritional, reducing amino acid absorption or dietary efficiency in human foods, especially if the seed is raw or undercooked. The proteins known to be problematic include lectins and trypsin inhibitors. Lectins are a class of glycoproteins that include alpha-amylases (AIs), arcelins (ARC), and phytohemagglutinin (PHA) that are useful in an evolutionary sense because they protect against insects and some abiotic stresses (drought, salinity); however they also inhibit food digestion of the globulins that make up the majority of seed protein (Mojica and de Mejía 2015). Trypsin inhibitors (TIs) are also known as serpins (serine protease inhibitors) that also inhibit the digestion of proteins.

Various protease inhibitors exist, but TIs form the largest and most diverse family of these potent anti-herbivory agents and are found in larger concentrations in wild accessions than in cultivated common beans (de Mejía et al. 2003). Fortunately, most of these problems can be overcome by cooking beans well past softness and eliminating the water used for soaking or cooking (Batista et al. 2010; Champ 2002). Eating raw beans or flours derived from them are never a good idea especially where dosage is not controlled. It is notable that unlike mung bean and other Asian legumes, few processed foods are made from common beans or their extracts and flours.

Phytates are also considered antinutrients by reducing the bioavailability of some proteins (Ariza-Nieto et al. 2007) and micronutrients to the human body (Huber 2016), although they are necessary for health and viability of the seedling common bean. As such, phytates are found principally in the cotyledon and embryo axis where they are immediately available to the growing plantlet that germinates from the common bean seed (Schlemmer et al. 2009). Total concentration of phytates can

be up to 3% of dry weight of common bean seeds (Blair et al. 2012). Phytates interact with some proteases to inhibit protein utilization in the plant or during digestion, so complex interactions of these factors must be considered when modifying phytate levels in common beans.

Phosphorus as a nutrient that makes up the bulk of the phytate molecules is needed at every growing stage of the plant and therefore is provided in abundance within legume seeds by the mother plants. Phytate concentration in seeds varies with the genotype selected for analysis, their efficiency at P use and mobilization, and the growing environment. Genotype x environmental effects are significant when experiments are done in the field before measuring phytate levels in common bean seeds. Phytates are usually higher in raw beans than in cooked ones so phytic acid can be reduced by cooking. Germinating or soaking the seed can also reduce phytates as these processes activate phytases. Breeding for lower phytates while considered a goal of some programs should be approached with caution as some low phytate mutants are less vigorous in production environments. In addition, some health benefits have been ascribed to phytates, and they play a role in slowing down amylase digest of starches.

Another aspect of phytates is that they are made up of various inositol types (IPs) with varying levels of phosphate groups, from IP1 to IP6. These different phytates vary in their ability to complex with Fe and Zn as micronutrients or Ca as an intermediate constitute of seeds (Rousseau et al. 2020a, 2020b). This influences the utility of common bean as a source of these minerals as well as P to the human diet. Mutants with low phytate production (*lpa*), especially for IP6 amounts, have been found in common bean by a program in Italy, and they have suggested that using this mutation may be a way to increase the bioavailability of Fe in biofortification of common bean (Campion et al. 2009; Cominelli et al. 2020).

Other potential antinutrients that also have roles as a nutraceutical is the complex group of flavonoids and tannins found in common bean seeds, primarily in the seed coat. Tannins are polyphenolics that can be condensed or uncondensed (Caldas and Blair 2009; Yang et al. 2018). As antinutrients, tannins inhibit absorption of various microelements (e.g., Fe) and also many proteins. They are generally rather heat stable and therefore remain after cooking affecting the color of the bean dish produced and how rapidly it oxidizes. Within the seed coat, tannins can also lead to darkening of the seed color so that light-colored beans such as cream mottled, cranberries, pinks, and pintos are often selected to be slow-darkening (Junk-Knievel et al. 2008). Flavonoids can be bioactive components and are found in many of the light-colored beans but not in white beans, while anthocyanins are found in black beans in particular (Moreno-Jiménez et al. 2019). These are water-soluble, and so in some common bean, cooking preparations can be reduced especially during long boiling. Cooking time can depend on seed hardness, seed coat thickness, and waxy seed coats that slow water absorption. Wax on the cuticle determine if the seed is shiny (now wax) or buff (wax present). Common bean tannins and flavonoids are probably less important in controlling disease compared to these components found in wine, where resveratrol is known to be a beneficial compound. The following table describes studies that have related common beans to health factors with an example given for each of the major human diseases affected by this pulse (Table 1).

Table 1 Research into common beans and human health for the discovery of nutraceuticals

Human health	Journal	Authors, year	Nutraceutical values
Coronary heart disease	*The American Journal of Clinical Nutrition, 74(4), 418–425.*	Nijveldt et al. (2001)	Quercetin has anti-inflammatory characteristics that blocks cyclooxygenase pathway and protects against coronary heart disease
	Official Journal of the Korean Physiological Society and the Korean Society of Pharmacology, 16(4), 249–253.	Jeon et al. (2012)	The Health Professionals Follow-Up Study found that males who followed a more sensible diet, which included more whole grains, legumes, fish, and chicken, had a 30% lower chance of developing heart disease
Antioxidant activity in common beans	*African Journal of Food, Agriculture, Nutrition and Development, 14(7), 9483–9496.*	Nyau (2014)	Methanol extract in the seed coat can stop oxidative chain reactions. Common beans have greatest antioxidant activity than 100 food items using in oxygen radical absorbance capacity (ORAC) assay
Cancer and Beans	*Agricultura técnicaen México, 33(1), 43–52*	Reynoso Camacho et al. (2007)	Dietary fiber may be protective against colorectal cancer. Pinto, black and white fed rats had four times less tumors than non-fed
Blood glucose level	*Legume Science, e155.*	Uebersax et al. (2022)	Legumes have moderate glycemic reactions
	Phytotherapy Research 16(7), 662–664.	Pari and Venkateswaran (2002)	*P. vulgaris* extracts have lower postprandial glycemia similar to metformin
Weight management	*Nutrition today, 39(1), 10–17*	Jones et al. (2004)	Diet with low glycemic index prevents onset of obesity

8 QTLs Discovered for Nutraceutical Breeding

Protein is the most cited nutritional component due to its often-cited high concentration of up to 22% in common bean (e.g., Jannat et al. 2019). The inheritance of protein content in common bean has been dominated by the dissection of the *Phs* locus for phaseolin type since this makes up over half the polypeptides in the seed (Viscarra-Torrico et al. 2021). This megalocus on chromosome Pv07 encodes the storage proteins, made up of many tandem copies of orthologous genes, which show up as the major QTL for protein concentration and protein quality (Joshi and Rao 2017;

Pandurangan and Marsolais 2017). Phaseolin is a type of globulin as mentioned earlier and is used for storing protein as the seed is developing and once it is fully grown (Fig. 5). Mutations that reduce the amount of the storage proteins increase the concentration of cysteine and methionine since these are not major components of the wild-type phaseolins (Taylor et al. 2008). Lectins follow phaseolins in abundance as seed proteins but are only 5–10 times as common. These proteins are also encoded at a megalocus that includes sequences for arcelin/phytohemagglutinin/α-amylase inhibitor (APA) on chromosome Pv04 (Viscarra-Torrico et al. 2021). QTLs for amino acid content are understudied except in the context of reduced phaseolin or lectin where cysteine and methionine contents increase, perhaps due to replacement of the two most common storage proteins by more unique proteins. This would be variable depending on the population studied (Table 2).

While carbohydrates are more abundant than protein at 60% or more of seed weight and of obvious organoleptic importance, it has been assumed to be quite static, which is unlikely given the large number of commercial classes of common bean and their variability in seed type. In addition, this is the portion of the seed most likely to change upon cooking. Pujolà et al. (2007) studied raw, soaked, and cooked states of eight Spanish beans versus a control Navy bean for amylose, amylopectin, digestible starch, and resistant starches finding that the latter two tend to increase in concentration with cooking if soaking water is removed. Of the other carbohydrate components, fiber and digest-resistant oligonucleotides have been the best studied. In the case of seed fiber, a pioneering study in the mapping of QTLs for fiber content in common bean was

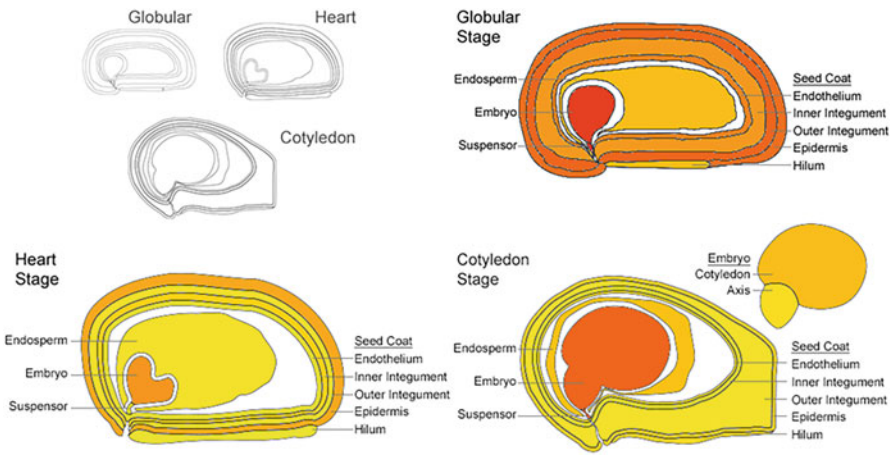


Fig. 5 Seed expression pattern for a phaseolin homolog globulin protein GLYMA03G399–40 in developing seed through three stages (heart, globular, and cotyledon) to show the intensity of tissue-specific levels of prevalence from low (yellow color) to medium (orange colors) to high (red). Subfigure at top left shows relative size of the developing seed at the three stages. Gene expression pattern for common bean is inferred from soybean as found in the eFP Browser (bar.utoronto.ca/efpsoybean/cgibin/efpWeb.cgi?dataSource=soybean_heart_cotyledon_globular) in which the protein is found early on in the seed coat, then in both seed coat and embryo and finally only in the embryo of *Glycine max*

Table 2 Germplasm and quantitative trait loci studies in common bean for nutritional components, a sampling

Title	Author	Journal	Population	Genotypes	Summary
Trypsin inhibitors, tannin, lectins	De Mejía et al. (2003)	*Journal of Agricultural and Food Chemistry, 51(20), 5962–5966*	5 bean cultivars	TI: 6.3–14.5 TUI/mg for 5 cultivars. Tannins:10.1–44.2 mg CE/g. Lectins: No significant diff.	The contribution of location to lectin of common bean was 13 times greater than cultivar and site interaction, implying that breeding for this feature may be difficult owing to a shortage of genetic resources For using beans as nutraceuticals, cultivars with high TI and tannin are promising because of their simple genetic control
Flavonoids	de Lima et al. (2014)	*Journal of Agricultural and Food Chemistry, 62(40), 9699–9704.*	16 genotypes	Kaempferol – highest in light color seed coat. Quercetin: 1.09–15.85 mg/g Myricetin: 26.15–78.94 mg/g	Significant quantities of flavonoids were found in common bean germplasm produced under the same environmental circumstances, with substantial diversity among groups with various seed coat colors
Proteins	Jannat et al. (2019)	*International Food Research Journal, 43(2), 595–601*	35 landraces from Pakistan +1 reference collection	Protein percent ranges from 18% to 22%	The common bean is the primary source of inexpensive protein for the people of many nations throughout the world. The protein profile revealed that the tested local bean ecotypes are effective protein sources for mountain dwellers
Minerals	Blair and Izquierdo (2012)	*International Journal of Plant Breeding Research, 62(40), 9699–9704*	7 pop's; 5 previously produced (2 Andean, 2 Middle American, 3 inter-genepool)	87 QTL (except QTL on chromosome 11) MQTL: 72 41 Andean, 25 Middle American, 6 wild Middle American	These findings demonstrated stronger consolidation than a maize meta-QTL study for grain Fe and Zn, which consolidated 28 unique QTL into 10 meta-QTL
Iron and zinc	Blair et al. (2010)	*Theoretical and Applied genetics, 121(6), 1059–1070.*	110 RILs in F_{10} generation developed by crossing G14519 × G4825	13 QTLs were identified for iron (5) and zinc (8)	A stable cross section QTL was found for seed mineral concentration associated with zinc and iron

Iron, zinc, phosphorous	Cichy et al. (2009)	*Crop Breeding & Genetic,49(5), 1742–1750*	77 RILs; AND696	Variation in seed Fe (40.8 mg/k) concentration in RIL was higher than Zn (16.5 mg/k). 46 QTLs are identified: 26% on linkage group B1 and more than1 trait affected; 28% linkage group B6 and multiple traits affected	Because there is substantial natural variety in seed Fe and Zn in common bean, such a goal is possible through plant breeding. Fe levels in the AND696/G19833 RIL population varied from 38 to 79 mg kg1, and Zn levels ranged from 16 to 33 mg kg1, all of which are within the predicted range based on bean germplasm screening
Anthocyanin	Díaz et al. (2010)	*Food Research International, 43(2), 595–601*	87 RILs	Anthocyanins: 0.013–0.21%. Most of the individuals did not pass 0.1%	Anthocyanins or polyphenolics are strong antioxidants that are thought to contribute to the health properties of beans
Macrominerals	Ribeiro et al. (2019)	*Genetics and Molecular Research, 18(2)*	100 RILs	Significant treatment effect is shown in potassium, calcium, and magnesium in F5:6 and F5:7 generations	Potassium-rich diets can lower blood pressure, especially in hypertensive people, and minimize the risk of diabetes. Calcium is required for the creation and maintenance of bones and skeletal structures. Magnesium is essential in the prevention of diabetes, osteoporosis, migraines, and cardiovascular disease
Folates	Khanal et al. (2013)	*International Journal of Agronomy, 1–9*	4 dry beans, 6F1 and 6F2	Total folate content: 217–345 in first injected solution; 167–231 mg/100 g in second injected solutions. g1268 (7.7%) and g2498 (7.8%) on Pv09 in first injection and g457_B in second injection shows significant dominant effect	The findings of this investigation indicated that, despite the prominent levels of folate concentration in dry beans, genetic diversity occurs between genotypes

carried out by Casanas et al. (2013) using a recombinant inbred line (RIL) population from the cross of a Spanish Alubia type Andean bean named Xana with a small black disease-resistant Mesoamerican bean named Cornell 49,242. They found a major QTLs associated with dietary fiber named DF7XC on chromosome Pv07, near the *p* locus involved in the genetic control of seed coat color.

Dietary fiber was also studied in a genome-wide association study (GWAS) panel of 280 cultivars from the Mesoamerican group of edible dry beans with 100 race Mesoamerica and 180 race Durango–Jalisco representatives by Moghaddam et al. (2018) who found a few more QTL for this trait. The same group also evaluated complex starches and found two candidate genes – *Phvul.001G214300* and *Phvul.001G215300* – for raffinose production that were homologous to Arabidopsis Stachyose Synthase (Sts)/Raffinose Synthase 4 (Rs4) and Galactinol Synthase 1 (GolS1), respectively.

Lipid content may vary between genotypes, but as this component is a small component of common bean, no current studies have looked at the inheritance of this trait. Some fat-soluble and insoluble vitamins have been evaluated. For example, Khanal et al. (2013) determined total folate content in four varieties of dry beans and their F_1 and $6F_2$ progeny finding that concentration could vary from 217 to 345 mg/100 g of seed in the first injected solution and from 167 to 231 mg/100 g in the second injected solutions with QTLs identified on chromosome Pv09 associated with markers g1268 (7.7% of variation explained) and g2498 (7.8%) for first injection and g457_B for second injection showing significant dominant effect. The findings of this investigation indicated that genetic diversity occurs between genotypes, despite the overall high levels of folate concentration in dry beans.

The seed coat is extremely important in the analysis of nutraceuticals as it is home to many of the health-benefiting minerals and secondary metabolites (Beninger and Hosfield 2003). Minerals are often bound to the tannins that accumulate in bean seed coats (Caldas and Blair 2009). Ca and to a lesser extent Mg are enriched in the seed coat (Aparicio-Fernandez et al. 2005).

Component of cell walls are compact in the hourglass and palisade layers that are external and internal facing cells. Two thirds to four fifths or more of seed Ca is found in the seed coat of common bean which contrast with soybean (Moraghan and Grafton 2001) and perhaps helps to explain differences in cooking time (Bassett et al. 2021). Tannins or polyphenolics accumulate at the seed coat cells that may also bind up $Fe^{2+/3+}$ and Zn^{2+} as well as Mg^{2+} as positively charged ions explaining their higher concentration.

Caldas and Blair (2009) analyzed three inter-genepool populations of RILs for condensed tannins from seeds and associated 12 QTLs with major genes for seed coat color. QTL explaining from 10% to 64% of the phenotypic variation for condensed tannin were found on linkage group B3, B7, and B10 associated with the Mendelian genes *Z*, *P*, and *Bip*, respectively. Most phenolics have strong antioxidant properties (Wang et al. 2016). Other than tannins, chief among the organic compounds found in seed coats are the flavonoids, flavones, flavonols, and isoflavonoids which include a range of polyphenolics from the flavonoid and isoflavonoid pathways (Fig. 6).

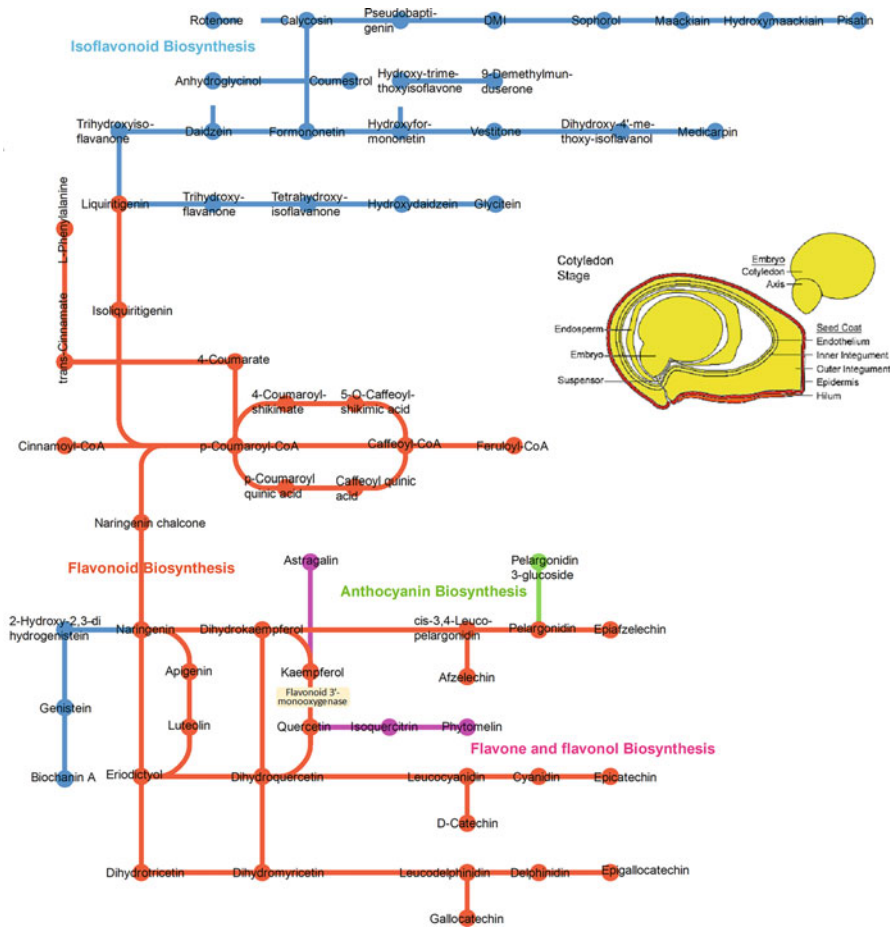


Fig. 6 A common bean specific drawing of the flavonoid (red), flavone/flavonol (pink), isoflavonoid (blue), and part of the anthocyanin (green) pathways with steps involved in intermediate and final products such as D-catechin, cyanidin, delphinidin, gallocatechin, kaempferol, pelargonidin, quercetin, etc. Subfigure to the right shows gene expression pattern for the highlighted gene flavonoid 3′ monooxygenase which converts kaempferol to quercetin. Gene expression pattern for common bean is inferred from soybean as found in the eFP Browser (bar.utoronto.ca/efpsoybean/cgibin/efpWeb.cgi?dataSource=soybean_heart_cotyledon_globular) in which the enzyme is produced only the epidermis of seed coat tissues both around the seed and in hilum areas of *Glycine max*. Pathway based on KEGG (http://www.genome.jp/kegg/) for genes expressed in *Phaseolus vulgaris* (pvu) using iPath v3.0 (https://pathways.embl.de/)

Authors have analyzed the flavonoids influencing seed color (Rocha-Guzmán et al. 2007). As a further example, de Lima et al. (2014) found that kaempferol was highest in light color seed coat among 16 genotypes evaluated. Quercetin ranged from 1.09 to 15.85 μg/g and myricetin from 26.15 to 78.94 μg/g. They concluded that significant quantities of flavonoids were found in common bean germplasm

produced under the same environmental circumstances, with substantial diversity among groups with various seed coat colors. Anthocyanins are an offshoot of the flavonoid pathway leading to pelargonidin and other dark colored red and purple tones, which are found mostly in dark red or small black beans. Díaz et al. (2010) used 87 RIL's from a combination of G19833 (sequenced genome) × DOR364 to find a range in anthocyanins of 0.013–0.21% skewed toward lower content. They suggested evaluation of anthocyanins and other polyphenolics as strong antioxidants that are thought to contribute to the health properties of beans.

The figure above highlights a key enzyme involved in metabolite synthesis for flavonoids, namely, the flavonoid 3′ monooxygenase for quercetin synthesis from kaempferol. This gene in *Glycine max* (soybean) was found with major expression in epidermis and hilum cells when KEGG and iPath were used to build a model of the metabolite profiles in the flavonoid pathway. It is important to consider that any of these phenolic derivatives have roles not only in seed coat color but also in anti-herbivory by insects or protection and camouflage from birds.

For phytates, another nutraceutical that has physiological roles in seed development and subsequent seedling growth, mutants have been analyzed to reduce the amount of phytates. Low phytic acid (*lpa*) mutants for the most part have abnormal plant phenotypes since phytic acid is a signaling molecular as well as phytates being a source of P for cell development. Blair et al. (2012) found QTL for phytate concentration in a central DOR364 × G19833 population that aligned with markers for the genes myo-inositol (3)P1 synthase (MIPS), myo-inositol kinase, and various inositol kinases. A phytate concentration QTL was found associated with one of two paralogs of the myo-inositol (3)P1 synthase gene family, located on linkage group Pv01 and expressed in common bean seed rather than in vegetative tissues. They concluded that natural variability in phytate levels was controlled by variation for MIPS activity as well as other loci that underlie the oligogenic inheritance in common bean. In Fig. 7, we show the primary enzymes and substrates/products in the inositol phosphate pathway, through the steps to phytates. One of the key enzymes is ATP1-myo-inositol 3′-phosphotransferase shown by KEGG mapping of annotated metabolites onto the pathway map from studies in many legumes (Darzi et al. 2018).

9 Current and Future Foci of Nutraceutical Research in Common Bean

To date, the best studied nutraceutical features of common bean have been protein, phytates, and several minerals considered important for biofortification, a process of improving legumes and other basic grains or starch sources for micronutrients (Blair 2013). Breeding for protein levels has had minimal success as most legumes already maximize their seed protein levels, and in common bean only a few mutants exist to change the makeup of those proteins that accumulate.

Meanwhile, phytates are important to biofortification with minerals because of chelation of iron, zinc, calcium, and other charged ions. Cominelli et al. (2020)

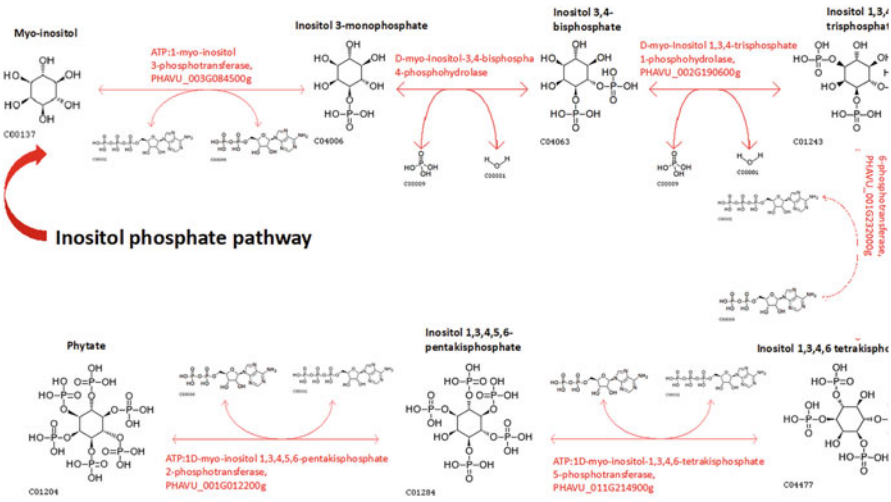


Fig. 7 The inositol phosphate pathway with enzymes (and gene accession IDin *Phaseolus vulgaris*) confirmed for common bean based on KEGG (http://www.genome.jp/kegg/) database

postulated that biofortification and plant adaptation may be at odds because of the inverse relationship between phytates and mineral bioavailability. Petry et al. (2010) found that this has real-world implications in feeding trials with common bean as a main source of iron for young women. Fiber also plays a complex role. Interactions during digestion including fermentation by microbes of fiber in the gut can affect human health. Butyrate, for example, is a product of fiber digestion that has anti-tumor and anti-inflammatory properties thus reducing cancer risk (Lanza et al. 2006; Chen et al. 2017).

Despite the interaction with tannins and phytates, the inheritance of mineral concentration has been a major focus of biofortification with the approach of "more is better" especially for iron and zinc which generally are deficient in grain-based or vegetarian diets (Blair 2013). The studies have been conducted in inter-genepool and within genepool populations. The former kind includes the pioneering study of Cichy et al. (2009) who found QTL for each of the two biofortification target mineral in the DOR364 × G19833 (Mesoamerican × Andean, $n = 87$) and AND696 × G19833 (Andean × Andean, $n = 77$) populations, respectively.

In the latter study, variation in seed Fe (40.8 mg/kg) concentration in RILs was higher than variation in Zn (16.5 mg/kg) with phytate also varying significantly resulting in a total of 46 QTLs found for iron, zinc, and phosphorus, the latter as a stand-in for phytates. In a third population, G14519 × G4825 (Mesoamerican × Mesoamerican) only 13 QTLs were found in total, 5 identified for Fe and 8 for Zn, but the population size was bigger ($n = 110$). A meta-analysis by Blair and Izquierdo (2012) looked at all previously produced populations and two new ones to identify 87 QTLs on every chromosome except Pv11. In more recent studies a greater number of minerals have been analyzed for QTL associations (Blair et al. 2016;

Ribeiro et al. 2019). While most of these studies looked at whole seed and did not attempt to breed lines with confirmed QTL, the studies by Blair et al. (2013a, b) were unique in using advanced backcrossing and seed coat analysis of minerals to do so. Advanced backcrossing was also used by Blair and Izquierdo (2012). Variability in national and international collections for biofortification traits has been confirmed by many authors (Celmeli et al. 2018; Delfini et al. 2021; Jan et al. 2021; Murube et al. 2021). Inheritance for all these micronutrients is mainly controlled as quantitative traits.

A more holistic approach is needed for true nutraceutical improvement of common bean. This should include analysis of nutrient levels as has been done with most studies conducted so far but also analysis of the genes involved in nutraceutical quality and physiological mechanisms leading to their expression. As an example of this, which admittedly is only an initial attempt to define some nutraceutical candidate genes, the KEGG pathway was used to target key enzymes in starch, protein storage, phytate synthesis, and flavonoid (anthocyanidin) synthesis pathway and find where in seed tissue the genes are expressed in an orthologous system.

To this end, we prepared a figure showing gene expression level in different seed compartments (Pelletier et al. 2017) for nine selected nutraceutical candidate genes (Fig. 8). In each case, the orthologous genes in soybean (*Glycine max*) were used to predict expression for common beans, as the eFP database does not yet include *Phaseolus vulgaris*. The 1,4-alpha-glucan branching enzyme (PHAVU_009G011000g), beta-amylase (PHAVU_011G107700g), and isoamylase (PHAVU_001G148700g) gene expression at globular seed stage indicate that alpha-

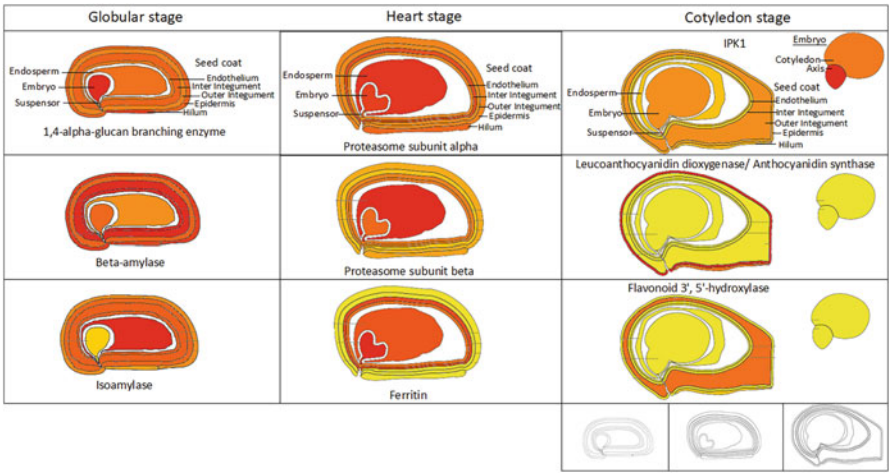


Fig. 8 Seed expression pattern for orthologs of nutrition candidate genes for common bean (*Phaseolus vulgaris*) based on RNA analysis of soybean (*Glycine max*) seeds where the intensity of tissue-specific levels of prevalence range from low (yellow color) to medium (orange colors) to high (red) based on the eFP Browser (bar.utoronto.ca/efpsoybean/). Subfigure at bottom in black and white shows relative size of the developing seed at the three stages

glucan enzyme has higher expressed level in the embryo. Isoamylase for starch degradation was mainly cotyledon expressed. Proteasome subunit alpha (PHAVU_008G093300g) and subunit beta (PHAVU_009G103400g) for protein degradation were highly expressed in heart stage cotyledons indicating places where phaseolin is accumulating and when.

Continuing with three more nutraceutical candidate genes: the enzyme highlighted for phytate accumulation. IPK1 (PHAVU_001G012200g) was observed to be highly expressed in the embryo axis at late stages. Finally, anthocyanidin synthase (PHAVU_002G152700g), an enzyme for anthocyanidin synthesis, was mainly expressed in seed coat epidermis. Similarly, flavonoid 3′,5′-hydroxylase (PHAVU_006G015400g) that converts quercetin to myricetin was mainly expressed in the seed coat outer integument but only at late stages of seed development. All these genes help researchers define where to look for nutraceutically important metabolites.

10 Conclusion

In conclusion, the common bean has been most often associated with diets of lower classes in developing countries, where it is known as "meat of the poor." This has created a bias toward studying only its basic proximate analysis of protein and starch contents for amino acid and caloric uptake or to evaluate its potential in biofortification for minerals such as Fe and Zn deficient in diets that are based on cereals. This reductionist attitude as exemplified in Pfeiffer and McClafferty (2007) does not holistically consider the full benefits that common beans can have on the human diet of both developed and developing countries.

Therefore, to reap the benefits of the entire suite of nutraceutical compounds in common beans, there is a need to evaluate the whole set of metabolites, their locations in the seed, and their interactions. This approach is practical for the proximate or reductionist researcher too. For example, during the biofortification project, phytates and tannins were not considered while breeding for higher minerals. And as a result, bioavailability has been ignored compared to total mineral accumulation, and the results have been mediocre advances in micronutrients that can be absorbed, in addition to problems with homeostasis of mineral content in the seed, considering where these are deposited. More reviews of all components of nutritional quality is needed since the biofortification aspect of common bean has been well reviewed (Blair 2013; Dwivedi et al. 2012) but other aspects have not. Here we have attempted to show the full benefits of common bean to human health through their nutraceutical properties. We have found that a lack of biochemical and functional studies stymies the advancement of common bean research in the legume nutraceutical space.

Acknowledgments The authors would like to thank the USDA Evans Allen program (TEN-X07) for funding this project as well as input from previous Bill and Melinda Gates Foundation projects. Useful discussions with previous and current students of the Blair Lab: Aura Diaz, Carolina

Astudillo, Daniel Ambachew, Gina Caldas, Karen Cichy, Josh Relyea, Max Miller, Maia Payne, Paulo Izquierdo, Tito Sandoval, and Tatiana Zabaleta, are also acknowledged. We also thank the USDA Nutrition laboratory in Ithaca New York under Dr. Ray Glahn for hosting many visits over the years, the ECABREN leadership under Dr. Paul Kimani, and the 1890s Capacity Building Program faculty development program for providing contexts in which to write this chapter.

References

Anderson JW, Major AW (2002) Pulses and lipaemia, short-and long-term effect: potential in the prevention of cardiovascular disease. Br J Nutr 88:S263–S271

Aparicio-Fernandez X, Yousef GG, Loarca-Pina G, De Mejia E, Lila MA (2005) Characterization of polyphenolics in the seed coat of Black Jamapa bean (*Phaseolus vulgaris* L.). J Agric Food Chem 53:4615–4622

Ariza-Nieto M, Blair MW, Welch RM, Glahn RP (2007) Screening of iron bioavailability patterns in eight bean (*Phaseolus vulgaris* L.) genotypes using the Caco-2 cell in vitro model. J Agric Food Chem 55:7950–7956

Bassett A, Hooper S, Cichy K (2021) Genetic variability of cooking time in dry beans (*Phaseolus vulgaris* L.) related to seed coat thickness and the cotyledon cell wall. Food Res Int 141:109886

Batista KA, Prudêncio SH, Fernandes KF (2010) Changes in the functional properties and anti-nutritional factors of extruded hard-to-cook common beans (*Phaseolus vulgaris*, L.). J Food Sci 75:C286–C290

Beninger CW, Hosfield GL (2003) Antioxidant activity of extracts, condensed tannin fractions, and pure flavonoids from *Phaseolus vulgaris* L. seed coat color genotypes. J Agric Food Chem 51: 7879–7883

Blair MW, Giraldo MC, Buendia HF, Tovar E, Duque MC, Beebe SE (2006) Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 113:100–109

Blair MW, Medina JI, Astudillo C, Rengifo J, Beebe SE, Machado G, Graham R (2010) QTL for seed iron and zinc concentration and content in a Mesoamerican common bean (*Phaseolus vulgaris* L.) population. Theor Appl Genet 121:1059–1070

Blair MW, Herrera AL, Sandoval TA, Caldas GV, Filleppi M, Sparvoli F (2012) Inheritance of seed phytate and phosphorus levels in common bean (*Phaseolus vulgaris* L.) and association with newly-mapped candidate genes. Mol Breed 30:1265–1277

Blair MW, Izquierdo P (2012) Use of the advanced backcross-QTL method to transfer seed mineral accumulation nutrition traits from wild to Andean cultivated common beans. Theor Appl Genet 125:1015–1031

Blair MW (2013) Mineral biofortification strategies for food staples: the example of common bean. J Agric Food Chem 61:8287–8294

Blair MW, Brondani RVP, Díaz LM, Del Peloso MJ (2013a) Diversity and population structure of common bean from Brazil. Crop Sci 53:1983–1993

Blair MW, Diaz LM, Acosta-Gallegos JA (2013b) Race structure in the Mexican collection of common bean landraces. Crop Sci 53:1517–1528

Blair MW, Wu X, Bhandari D, Astudillo C (2016) Genetic dissection of ICP-detected nutrient accumulation in the whole seed of common bean (*Phaseolus vulgaris* L.). Front Plant Sci 7:219

Caldas GV, Blair MW (2009) Inheritance of seed condensed tannins and their relationship with seed-coat color and pattern genes in common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 119:131–142

Campion B, Sparvoli F, Doria E, Tagliabue G, Galasso I, Fileppi M, Bollini R, Nielsen E (2009) Isolation and characterisation of an *lpa* (low phytic acid) mutant in common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 118:1211–1221

Carbonaro M (2006) 7S globulins from *Phaseolus vulgaris* L.: impact of structural aspects on the nutritional quality. Biosci Biotechnol Biochem 70:2620–2626

Carbonaro M, Maselli P, Nucara A (2015) Structural aspects of legume proteins and nutraceutical properties. Food Res Int 76:19–30

Casanas F, Pérez-Vega E, Almirall A, Plans M, Sabaté J, Ferreira JJ (2013) Mapping of QTL associated with seed chemical content in a RIL population of common bean (*Phaseolus vulgaris* L.). Euphytica 192:279–288

Celmeli T, Sari H, Canci H, Sari D, Adak A, Eker T, Toker C (2018) The nutritional content of common bean (*Phaseolus vulgaris* L.) landraces in comparison to modern varieties. Agronomy 8:166

Champ MMJ (2002) Non-nutrient bioactive substances of pulses. Br J Nutr 88:S307–S319

Chen PX, Zhang H, Marcone MF, Pauls KP, Liu R, Tang Y, Zhang B, Renaud JB, Tsao R (2017) Anti-inflammatory effects of phenolic-rich cranberry bean (*Phaseolus vulgaris* L.) extracts and enhanced cellular antioxidant enzyme activities in Caco-2 cells. J Funct Foods 38:675–685

Cichy KA, Caldas GV, Snapp SS, Blair MW (2009) QTL analysis of seed iron, zinc, and phosphorus levels in an Andean bean population. Crop Sci 49:1742–1750

Cominelli E, Pilu R, Sparvoli F (2020) Phytic acid and transporters: what can we learn from low phytic acid mutants? Plan Theory 9:69

Cvitanich C, Przybyłowicz WJ, Urbanski DF, Jurkiewicz AM, Mesjasz-Przybyłowicz J, Blair MW, Astudillo C, Jensen EØ, Stougaard J (2010) Iron and ferritin accumulate in separate cellular locations in Phaseolus seeds. BMC Plant Biol 10:1–14

Darzi Y, Letunic I, Bork P, Yamada T (2018) iPath3. 0: interactive pathways explorer v3. Nucleic Acids Res 46:W510–W513

de Lima PF, Colombo CA, Chiorato AF, Yamaguchi LF, Kato MJ, SrAM C (2014) Occurrence of isoflavonoids in Brazilian common bean germplasm (*Phaseolus vulgaris* L.). J Agric Food Chem 62:9699–9704

De Mejía EG, Guzmán-Maldonado SH, Acosta-Gallegos JA, Reynoso-Camacho R, Ramírez-Rodríguez E, Pons-Hernández JL, González-Chavira MM, Castellanos JZ, Kelly JD (2003) Effect of cultivar and growing location on the trypsin inhibitors, tannins, and lectins of common beans (*Phaseolus vulgaris* L.) grown in the semiarid highlands of Mexico. J Agric Food Chem 51:5962–5966

Delfini J, Moda-Cirino V, dos Santos NJ, Zeffa DM, Nogueira AF, Ribeiro LAB, Ruas PM, Gepts P, Gonçalves LSA (2021) Genome-wide association study for grain mineral content in a Brazilian common bean diversity panel. Theor Appl Genet 134:2795–2811

Dhital S, Bhattarai RR, Gorham J, Gidley MJ (2016) Intactness of cell wall structure controls the in vitro digestion of starch in legumes. Food Funct 7:1367–1379

Díaz AM, Caldas GV, Blair MW (2010) Concentrations of condensed tannins and anthocyanins in common bean seed coats. Food Res Int 43:595–601

Díaz LM, Blair MW (2006) Race structure within the Mesoamerican gene pool of common bean (*Phaseolus vulgaris* L.) as determined by microsatellite markers. Theor Appl Genet 114:143–154

Dwivedi SL, Sahrawat KL, Rai KN, Blair MW, Andersson MS, Pfeiffer W (2012) Nutritionally enhanced staple food crops. Plant Breed Rev 36:169–292

Emani C, Hall TC (2008) Phaseolin: structure and evolution. Open Evol J 2:66

Florez A, Pujolà M, Valero J, Centelles E, Almirall A, Casañas F (2009) Genetic and environmental effects on chemical composition related to sensory traits in common beans (*Phaseolus vulgaris* L.). Food Chem 113:950–956

Gomes F, Ferreira A, Wojeicchowski JP, Nogueira A, Demiate IM (2018) Beans (*Phaseolus vulgaris* L.): whole seeds with complex chemical composition. Curr Opin Food Sci 19:63–71

Hayat I, Ahmad A, Masud T, Ahmed A, Bashir S (2014) Nutritional and health perspectives of beans (*Phaseolus vulgaris* L.): an overview. Crit Rev Food Sci Nutr 54:580–592

Huber K (2016) Phenolic acid, flavonoids and antioxidant activity of common brown beans (*Phaseolus vulgaris* L.) before and after cooking. J Nutr Food Sci 06(05)

Izquierdo P, Astudillo C, Blair MW, Iqbal AM, Raatz B, Cichy KA (2018) Meta-QTL analysis of seed iron and zinc concentration and content in common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 131:1645–1658

Jan S, Rather IA, Sofi PA, Wani MA, Sheikh FA, Bhat MA, Mir RR (2021) Characterization of common bean (*Phaseolus vulgaris* L.) germplasm for morphological and seed nutrient traits from Western Himalayas. Legume Sci 3:e86

Jannat S, Shah AH, Sabir SM (2019) Nutraceutical characterisation of common bean (*Phaseolus vulgaris* L.) germplasm from Pakistan. Int Food Res J 26:1835–1843

Jeon S, Han S, Lee J, Hong T, Yim D-S (2012) The safety and pharmacokinetics of cyanidin-3-glucoside after 2-week administration of black bean seed coat extract in healthy subjects. Korean J Physiol Pharmacol 16:249–253

Jones JM, Reicks M, Adams J, Fulcher G, Marquart L (2004) Becoming proactive with the whole-grains message. Nutr Today 39:10–17

Joshi PK, Rao PP (2017) Global pulses scenario: status and outlook. Ann N Y Acad Sci 1392:6–17

Junk-Knievel DC, Vandenberg A, Bett KE (2008) Slow darkening in pinto bean (*Phaseolus vulgaris* L.) seed coats is controlled by a single major gene. Crop Sci 48:189–193

Khanal S, Xue J, Khanal R, Xie W, Shi J, Pauls KP, Navabi A (2013) Quantitative trait loci analysis of folate content in dry beans, *Phaseolus vulgaris* L. Int J Agron 2013:1–9

Lanza E, Hartman TJ, Albert PS, Shields R, Slattery M, Caan B, Paskett E, Iber F, Kikendall JW, Lance P (2006) High dry bean intake and reduced risk of advanced colorectal adenoma recurrence among participants in the polyp prevention trial. J Nutr 136:1896–1903

Ma M, Wang Y, Wang M, Jane J-l, Du S-k (2017) Physicochemical properties and in vitro digestibility of legume starches. Food Hydrocoll 63:249–255

Moghaddam SM, Brick MA, Echeverria D, Thompson HJ, Brick LA, Lee R, Mamidi S, McClean PE (2018) Genetic architecture of dietary fiber and oligosaccharide content in a middle american panel of edible dry bean. Plant Genome 11:170074

Mojica L, de Mejía EG (2015) Characterization and comparison of protein and peptide profiles and their biological activities of improved common bean cultivars (*Phaseolus vulgaris* L.) from Mexico and Brazil. Plant Foods Hum Nutr 70:105–112

Mojica L, de Mejia EG, Granados-Silvestre MÁ, Menjivar M (2017) Evaluation of the hypoglycemic potential of a black bean hydrolyzed protein isolate and its pure peptides using in silico, in vitro and in vivo approaches. J Funct Foods 31:274–286

Moraghan JT, Grafton K (2001) Genetic diversity and mineral composition of common bean seed. J Sci Food Agric 81:404–408

Moraghan JT, Grafton K (2002) Distribution of selected elements between the seed coat and embryo of two black bean cultivars. J Plant Nutr 25:169–176

Moreno-Jiménez MR, López-Barraza R, Cervantes-Cardoza V, Pérez-Ramírez IF, Reyna-Rojas JA, Gallegos-Infante JA, Estrella I, Rojas-Contreras JA, González-Laredo RF, Rocha-Guzmán NE (2019) Mechanisms associated to apoptosis of cancer cells by phenolic extracts from two canned common beans varieties (*Phaseolus vulgaris* L.). J Food Biochem 43:e12680

Murube E, Beleggia R, Pacetti D, Nartea A, Frascarelli G, Lanzavecchia G, Bellucci E, Nanni L, Gioia T, Marciello U (2021) Characterization of nutritional quality traits of a common bean germplasm collection. Foods 10:1572

Nijveldt RJ, Van Nood EL, Van Hoorn DE, Boelens PG, Van Norren K, Van Leeuwen PA (2001) Flavonoids: a review of probable mechanisms of action and potential applications. Am J Clin Nutr 74(4):418–425

Nyau V (2014) Nutraceutical perspectives and utilization of common beans (*Phaseolus vulgaris* L.): a review. Afr J Food Agric Nutr Dev 14:9483–9496

OECD (2019) Safety assessment of foods and feeds derived from transgenic crops, volume 3: common bean rice cowpea and apple compositional considerations novel food and feed safety. OECD Publishing, Paris

Osborn TC, Hartweck LM, Harmsen RH, Vogelzang RD, Kmiecik KA, Bliss FA (2003) Registration of *Phaseolus vulgaris* genetic stocks with altered seed protein compositions. (Registrations of genetic stocks). Crop Sci 43:1570–1572

Pandurangan S, Marsolais F (2017) Analysis of phaseolin copy number by quantitative PCR. Annu Rep Bean Improv Coop 60:107–108

Pari L, Venkateswaran S (2002) Hypoglycaemic activity of Scopariadulcis L. extract in alloxan induced hyperglycaemic rats. Phytother Res 16:662–664

Peláez D, Aguilar PA, Mercado M, López-Hernández F, Guzmán M, Burbano-Erazo E, Denning-James K, Medina CI, Blair MW, De Vega JJ (2022) Genotype selection, and seed uniformity and multiplication to ensure common bean (*Phaseolus vulgaris* L.) var. Liborino Agron 12:2285

Pelletier JM, Kwong RW, Park S, Le BH, Baden R, Cagliari A, Hashimoto M, Munoz MD, Fischer RL, Goldberg RB (2017) LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. PNAS 114:E6710–E6719

Petry N, Egli I, Zeder C, Walczyk T, Hurrell R (2010) Polyphenols and phytic acid contribute to the low iron bioavailability from common beans in young women. J Nutr 140:1977–1982

Pfeiffer WH, McClafferty B (2007) HarvestPlus: breeding crops for better nutrition. Crop Sci 47:S-88

Pujolà M, Farreras A, Casañas F (2007) Protein and starch content of raw, soaked and cooked beans (*Phaseolus vulgaris* L.). Food Chem 102:1034–1041

Rebello CJ, Greenway FL, Finley JW (2014) A review of the nutritional value of legumes and their effects on obesity and its related co-morbidities. Obes Rev 15:392–407

Reynoso Camacho R, Ríos Ugalde MD, Torres Pacheco I, Acosta Gallegos JA, Palomino Salinas AC, Ramos Gómez M, González Jasso E, Guzmán Maldonado SH (2007) Common bean (Phaseolus vulgaris L.) consumption and its effects on colon cancer in Sprague-Dawley rats. Agricultura Técnica en México 33(1):43–52

Ribeiro ND, Mezzomo HC, Santos GG (2019) Genetic parameters and combined selection for seed coat color and macrominerals in Mesoamerican common bean lines. Gen Mol Res 18:1–14

Rivera A, Plans M, Sabaté J, Casañas F, Casals J, Rull A, Simó J (2018) The Spanish core collection of common beans (*Phaseolus vulgaris* L.): an important source of variability for breeding chemical composition. Front Plant Sci 9:1642

Rocha-Guzmán NE, Herzog A, González-Laredo RF, Ibarra-Pérez FJ, Zambrano-Galván G, Gallegos-Infante JA (2007) Antioxidant and antimutagenic activity of phenolic compounds in three different colour groups of common bean cultivars (*Phaseolus vulgaris*). Food Chem 103: 521–527

Rodiño AP, Santalla M, De Ron AM, Singh SP (2003) A core collection of common bean from the Iberian peninsula. Euphytica 131:165–175

Rousseau S, Celus M, Duijsens D, Gwala S, Hendrickx M, Grauwet T (2020a) The impact of postharvest storage and cooking time on mineral bioaccessibility in common beans. Food Funct 11:7584–7595

Rousseau S, Pallares AP, Vancoillie F, Hendrickx M, Grauwet T (2020b) Pectin and phytic acid reduce mineral bioaccessibility in cooked common bean cotyledons regardless of cell wall integrity. Food Res Int 137:109685

Sathe SK, Deshpande SS, Salunkhe DK, Rackis JJ (1984) Dry beans of Phaseolus. A review. Part 2. Chemical composition: carbohydrates, fiber, minerals, vitamins, and lipids. Crit Rev Food Sci Nutr 21:41–93

Schlemmer U, Frølich W, Prieto RM, Grases F (2009) Phytate in foods and significance for humans: food sources, intake, processing, bioavailability, protective role and analysis. Mol Nutr Food Res 53:S330–S375

Singh B, Singh JP, Shevkani K, Singh N, Kaur A (2017) Bioactive constituents in pulses and their health benefits. J Food Sci Technol 54:858–870

Taylor M, Chapman R, Beyaert R, Hernández-Sebastià C, Fdr M (2008) Seed storage protein deficiency improves sulfur amino acid content in common bean (*Phaseolus vulgaris* L.): redirection of sulfur from γ-glutamyl-S-methyl-cysteine. J Agric Food Chem 56:5647–5654

Uebersax MA, Cichy KA, Gomez FE, Porch TG, Heitholt J, Osorno JM, Kamfwa K, Snapp SS, Bales S (2022) Dry beans (*Phaseolus vulgaris* L.) as a vital component of sustainable agriculture and food security—a review. Legume Sci:e155

Viscarra-Torrico RC, Pajak A, Garzón AS, Zhang B, Pandurangan S, Diapari M, Song Q, Conner RL, House JD, Miklas PN (2021) Common bean (*Phaseolus vulgaris* L.) with increased cysteine and methionine concentration. Legume Sci 3:e103

Wang N, Toews R (2011) Certain physicochemical and functional properties of fibre fractions from pulses. Food Res Int 44:2515–2523

Wang Y-K, Zhang X, Chen G-L, Yu J, Yang L-Q, Gao Y-Q (2016) Antioxidant property and their free, soluble conjugate and insoluble-bound phenolic contents in selected beans. J Funct Foods 24:359–372

Yadav BS, Sharma A, Yadav RB (2010) Resistant starch content of conventionally boiled and pressure-cooked cereals, legumes and tubers. J Food Technol Res 47:84–88

Yang QQ, Gan RY, Ge YY, Zhang D, Corke H (2018) Polyphenols in common beans (*Phaseolus vulgaris* L.): chemistry, analysis, and factors affecting composition. Compr Rev Food Sci Food Saf 17:1518–1539

Genetic Improvement of Nutraceutical Traits in Chickpea (*Cicer arietinum* L.)

Alok Das, Biswajit Mondol, Prateek Singh, and Shallu Thakur

Contents

Abstract

Chickpea is a rich source of essential dietary nutrients and is considered as a functional food that offers various health benefits, including maintaining healthy gut, reduce obesity, aid in weight management, etc. Chickpea plays a vital role in ensuring nutritional food security and is considered as chief source of protein, carbohydrate, fats, fibers, folate, beta carotene, and different elements. Beside nutritional significance chickpea has the therapeutic properties and is claimed to inhibit cancer and is involved in maintaining normal blood pressure and assists in regulating type-2 diabetes and various cardiac disease. Apart from these nutritional benefits, chickpeas have been reported for the presence of some antinutritional factors (ANFs) that can be minimized employing various processing techniques. Efforts are in progress to identify QTLs linked to enhanced nutritional content and

A. Das (✉) · B. Mondol · P. Singh · S. Thakur
Division of Plant Biotechnology, ICAR-Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India
e-mail: alok.das@icar.gov.in

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_24

genome wide association of protein, beta-carotene, Fe, Zn, Cu, P, and K in a panel of chickpea genotypes. Biotechnological interventions for biofortification of chickpea included combination of chickpea nicotianamine synthase 2 (*CaNAS2*) and soybean ferritin (*GmFER*) genes that have reported to enhance iron content from the transgenic seeds, however, further study is required in this area. Chickpea is widely adopted in various cuisines and forms across populations, and its nutritional enrichment should be the priority area of research.

Keywords

Protein-rich legumes · Nutraceutical duality · Nutritional diversity · Mapping · QTLs · Genome wide association studies (GWAS) · Genetic engineering

1 Introduction

Chickpea is a grain legume belonging to the *Leguminosae* family that are prominently cultivated in Asia, Africa, the Middle East, Mediterranean Europe, Canada, Australia, and some parts of South America. The genus *Cicer* comprises of one cultivated species (*Cicer arietinum* L.) and 42 wild species. Chickpea originated from South East Turkey in the Fertile Crescent belt near Syria. Four centers of diversity were identified, viz., Mediterranean, Central Asia, the Near East, and India, besides, the secondary center of origin in Ethiopia (Vavilov 1951). Van der Maesen (1972) reported that the species originated in the southern Caucasus and northern Persia. Later, Ladizinsky (1975) reported southeastern Turkey as the center of origin. Based on the presence of the closely related annual *Cicer* species, *C. reticulatum,* and *C. echinospermum*, P.H. Davis, Van der Maesen (1987) opined the southeastern part of Turkey adjoining Syria as the possible center of origin. Wild *C. reticulatum* is intercrossable with the cultivated species *C. arietinum* and resembles morphologically as well. It was regarded as the wild progenitor of cultivated species, *C. arietinum* (Ladizinsky 1975). Duke 1981 also reported chickpeas were first domesticated in the Middle East and were widely cultivated in India, Mediterranean area, the Middle East, and Ethiopia. Wild *Cicer* species were reported to be abundant in Turkey, Iran, Afghanistan, and Central Asia (Duke 1981). The crop subsequently spread with human migration toward the West and South via the Silk Route (Singh et al. 1997). Chickpea is diploid and primarily self-pollinated (occasional cross-pollination by insects up to 1%). Thirty-six of the wild relatives are perennials and the ten are annual (Toker et al. 2021). It is the world's second largest legume crop based on the annual production and India being the largest producer accounting for the two-third of the total production (FAOSTAT 2020). Chickpea are broadly classified into two major classes: *Desi* and *Kabuli* based on the seed morphology. The *Desi* variety have relatively smaller seeds with dark colored thick seed coat whereas, the *Kabuli* variety have pale cream-colored thinner seed coat.

Chickpea plays a vital role in the diet of humans and animals and also helps to maintaining soil fertility, especially in arid areas. The ability of chickpea to fix

nitrogen in the soil makes it an important choice in farming systems. Under favorable environmental condition, chickpea can fix up to 60–80 kg/ha of atmospheric nitrogen. The symbiosis process in *Rhizobia* is governed by mainly two classes of genes involved in the nodulation process and other employed in nitrogen fixation. The nodulation genes are important component of the rhizobium-legume pairing that are involved in the production of enzymes that play key role in the biosynthesis and secretion of Nod factors that interact with plant flavonoids for initiating nodulation (Via et al. 2016). The genes responsible for nitrogen fixation include nitrogenase enzyme genes that are involved in conversion of free atmospheric nitrogen into ammonia. Chickpea are vulnerable to climate change, both drought and heat can significantly reduce its productivity. Temperatures and drought tolerance are the key traits in chickpea breeding for maintaining the production of chickpea in adverse areas.

In the contemporary era, with the change in lifestyle and food habits, nutritional quality has been a matter of consideration. Pulses together share only 5% of total protein consumption globally (Joshi and Parthasarathy Rao 2016), but it is significantly higher in India (6–24%) (NNMB 2006). Legumes have been scored >100 along with animal-based proteins for amino acid score based on lysine content (WHO 2007), which were conventionally grouped as *second-class* proteins by dieticians with respect to the animal-based protein (classified as *first-class* protein). Animal protein-based food contains cholesterol which is not advantageous for the human compared to plant-based protein foods. Chickpea is the major food legume or pulse crop in South Asia and it figures in the dietary habits of people in Africa, Mediterranean Europe, South and North America as well. Owing to its nutritional quality, chickpeas are currently being tested in space programs.

2 Nutritional Composition

Chickpea, also known as *garbanzo beans*, has been part of human diet since ancient times. Its leaves and seeds are globally consumed in several different forms and preparations based on the ethnic and regional factors. Chickpea has been the part of staple diet among the population living in the Middle East, India, Israel, Spain, and North Africa. The soaked chickpea seed salad is a healthy alternative for weight management, the dehusked split seeds are consumed as *dal,* and the whole seeds are consumed in boiled, roasted, and fermented forms in the Indian subcontinent.

Chickpea seeds contains carbohydrates (50–58%), protein (15–22%), moisture (7–8%), fat (3.8–10.20%), and micronutrients (<1%) (USDA 2021). They are also rich in globulin (42%), albumin (16%), and glutens (10%). They contain all essential dietary amino acids except the sulphur-containing amino acids that can be complemented from cereals to form a balanced diet when consumed together. Chickpeas are rich in carbohydrates containing both simple sugars (glucose, fructose, and ribose) and complex sugars (raffinose, ciceritol, and starch). Besides nutritional benefits, chickpea also contains prebiotics that have been claimed to provide several health benefits including maintaining healthy gut, reduce obesity,

and aid in weight management. Furthermore, prebiotic carbohydrates assists in lipid metabolism producing short chain fatty acids that are linked with various health benefits including reducing obesity, reduces insulin dependency, and prevents colorectal cancer. Prebiotics carbohydrates are complex sugars that include simple sugars, fructo-oligosaccharides (ketose and nystose), inulin, resistant starch, sugar alcohols (sorbitol and mannitol), and raffinose oligosaccharides (raffinose, stachyose, and verbacose). Chickpea contains various carotenoids and vitamins including β-carotenoids, canthaxanthin, xanthophyll, tocopherols, and vitamin B complex (B2, B5, B6, and B9).

Chickpea being a nonoil seed crop has low fat content (3–10%) but is higher compared to other pulses. Major classes of fatty acids found in chickpea originated from storage lipids (triacylglycerols) (Jukanti et al. 2012). Sterols and phytosterols and various classes of polyunsaturated fatty acids (66%) including essential fatty acids (ω-6 and ω-3), monounsaturated fatty acids (19%), and saturated fatty acids (15%) are present in chickpea. In general, chickpea seeds have higher fat content than most of the pulses and cereals consumed globally. Although chickpea has low lipid content, but the two prominent polyunsaturated fatty acids (PUFAs) include linoleic and oleic acids that constitute more than 50% of total chickpea lipid reserve. PUFAs belong to the class of good cholesterols, an essential element that is required to regulate the level of bad cholesterol in humans.

Chickpea seeds and sprouts are rich source flavonoids, especially isoflavones. Carotenoids are brightly colored lipid-soluble pigments that exhibit antioxidants activity. Carotenoids are of two types: oxygenated type (xanthophylls that includes lutein, violaxanthin, and neoxanthin), and the nonoxygenated type (carotenes that includes β-carotene and lycopene). The major classes of carotenoids present in chickpea include α-carotene, β-carotene, lutein, zeaxanthin, lycopene, and β-crypto xanthin. Interestingly, average concentration of carotenoids present in the wild accessions of chickpea is higher than the cultivated varieties. Among all the carotenoids found in plants, the β-carotene is the most widely distributed carotenoid that can be converted into vitamin A efficiently (Abbo et al. 2005).

Chickpeas are rich source of dietary minerals like calcium, zinc, phosphorus, magnesium, and folate. Furthermore, it is a chief source of other vitamins including riboflavin, niacin, thiamin, and folate. Chickpea serves as a major source of protein and carbohydrate among the vegetarian population living in the semi-arid tropics. They are cheap source of dietary protein and also sometimes referred as “poor man’s meat.”

3 Medicinal and Physiological Properties

Isoflavones, linoleic acid, saponin, and β-sitosterol found in chickpea exhibit therapeutic significance as they are associated with anticancerous properties and assist in cholesterol management and promote cardiac health. Isoflavones found in chickpea have been reported to inhibit development of breast cancer cells by initiating apoptosis of the MCF-7 cells at a concentration of 32.8 μg/ml. Inhibition of cell

proliferation is governed by up-regulation of apoptosis signaling pathway and down-regulation of genes involved in mRNA splicing pathway. The modified expression of 11 hub genes of PPI (protein-protein interaction) networks under the effect of isoflavones enhances the survival time and is potential anticancer agent in breast cancer patients (Wang et al. 2020). Chickpea consumption had been associated with a reduction in the incidence of HDL cholesterol, type-2 diabetes, and heart disease. Linoleic acid, a dominant PUFA which is involved in maintaining normal blood pressure and its involvement in production of prostaglandins had been reported in chickpea. Saponin rich foods have been reported to reduce plasma cholesterol significantly (Singh et al. 2017). Saponin binds with the dietary cholesterol, thereby promoting their egestion through faeces. Saponins have also been reported with strong cytotoxic effects against cancer cell lines. β-sitosterol (dominant phytosterol in chickpea) have been reported effective in reducing serum cholesterol levels and incidence of coronary heart disease. Recommended intake of folic acid had been reported to lower the serum homocysteine concentrations, a risk factor for chronic heart disease. It is found that people with recommended fibre intakes (26 g/day) had a 27% lower risk of cardiovascular diseases compared to those with low intakes (12 g/day). Deposition of visceral and ecotopic fats, resulting in hypolipidaemia and insulin-sensitizing effects in the rats were reduced by inclusion of chickpea in high-fat rodent feed.

Chickpea and yellow pea hydrolysates produced by gastrointestinal enzymes had high ACE inhibitory activity which was comparable to milk protein hydrolysates. Furthermore, peptides derived from garden pea, cowpea, and chickpea exhibits antioxidant, hypo-cholesterolemic, anticancer, and antimicrobial properties. The chickpea derived peptides had been reported to inhibit the growth of Caco-2 and THP-1 cell lines. The cancerous cell lines Caco-2 responsible for tumors in the colon and THP-1 cells grow in basal space. The antioxidant and free radical scavenging activities of chickpea protein hydrolysates (CPH). The CPHs were assayed based on reducing power, inhibition of linoleic acid autoxidation, and 1,1-diphenyl-2-picrylhydrazyl (DPPH)/ superoxide/ hydroxyl radical scavenging assays. The authors reported antioxidant activity of lowest molecular weight fraction as 81.13%, which was closer to that of α-tocopherol (83.66%) in the linoleic acid oxidation system.

4 Nutraceutical Duality

Food legumes contain several antinutritional factors (ANFs) like protease inhibitors, amylase inhibitors, lectins, saponins, phytic acid, and long chain oligosaccharides (raffinose family)/ polysaccharides. Biochemistry research of 1980s and 1990s concentrated majorly on quantification and characterization of those ANFs in food legume crops. Singh (1988) reported various antinutritional factors present in chickpea, viz., protease inhibitors, amylase inhibitors, phytolectins, polyphenols, and oligosaccarides that can be significantly reduced through processing practices as cooking, germination, and fermentation. Mittal et al. (2012) studied different processing treatments, namely,

germination, boiling, pressure cooking, and roasting for reduction of various antinutritional factors like phytic acid, polyphenols, tannins, saponins, oxalates, and trypsin inhibitor activity in chickpea. The authors reported pressure cooking as maximum effective methods in reduction of all types of antinutritional factors. Pressure cooking reduced tannins by 93.97% and polyphenols by 87.71% which were maxima. Protein fraction, fatty acid profile, and mineral content of chickpea were affected by processing treatments. Notably, the albumin fraction was least affected on processing. However, germination increased the linolenic acid to 48.42%, Fe to 56.89%, and K to 28.6% respectively. Cooking treatments observed to be causing significant reduction in fat, total ash, carbohydrate fractions, antinutritional factors (trypsin inhibitor, haemagglutinin activity, tannins, saponins, and phytic acid), minerals, and water-soluble B-vitamins. Significant reduction was observed after cooking treatment for reducing sugars, sucrose, trisaccharides (raffinose and stachyose); whereas verbascose was completely eliminated. Authors reported cooking treatments affect protein profile by reducing the concentrations of lysine, tryptophan, and total aromatic amino acids. However, after the reduction in concentration, cooked chickpeas were still higher in lysine, isoleucine, and total aromatic amino acid contents with respect to the standard (FAO and WHO). The authors reported lesser reduction in B-vitamins and minerals in chickpeas cooked by microwaving than those cooked by boiling and autoclaving. Chickpeas subjected to the various cooking treatments varied considerably, depending on the type of treatment. Based on the chemical score and limiting amino acid, the authors recommended microwave cooking for chickpea preparation, which not only improved nutritional quality (by reducing the level of antinutritional and flatulence factors, increasing in-vitro protein digestibility and retention rates of both B-vitamins and minerals) but also reduced the cooking time. Singh et al. (2017) reviewed illustratively regarding saponins in chickpea, where processing and cooking have been reported to reduce them. The effect of soaking, sprouting, and cooking on the stability and bioavailability of saponins has been studied to limited extent.

The duality regarding the ANFs is interesting and intriguing. ANF like raffinose family of oligosaccharides (RFO) which consist of oligosaccharides like raffinose, stachyose, verbascose are commonly found in pulse crops and often associated with flatulence. However, recent researches suggest RFOs reduce the risk of intestinal cancer and promote bowel clearance. During germination, RFOs serve as a carbon source. The long chain oligosaccharides or polysaccharides help in reducing the glycemic index by reducing the rate of digestion. It reduces hankering for carbohydrates and is utilized by dieticians for weight management diets. Another ANF, haemagglutinogen exhibit beneficial role in arresting metastasis and killing cancerous cells, activation of body defense mechanisms along with obesity reduction. Saponins are another group of ANF present in legume crops which has been reported to trigger RBC hemolysis in higher concentration. Recent researches report saponins to decrease the serum lipid and glucose level and also lower cancer risk. Diets rich in saponin have been reported to prevent dental carries as well. Amylose inhibitors have potential to be used in treatment strategy to control diabetes for its property to reduce the digestion of starch and altar response of sugar to insulin. Hemolytic disorder causing ANFs from fababean, viz., vicine and convicine have also been

reported to reduce cholesterol. The beneficial attributes of protease inhibitor (ANFs), viz., trypsin and chymotrypsin inhibitors include their anticancerous, anti-inflammatory properties. The authors also reported ANF Angiotensin I-converting enzyme (ACE) inhibitor to be associated with control of hypertension. Replacement of animal protein foods with legume proteins limited the intake of saturated fats and increased the intake of dietary fibers, which helped lowering the risk of cardiovascular disorders. WHO in its 2008 report has stated that adoption of pulse-based diets prevents up to 80% of heart disease, stroke, type 2 diabetes, and some cancers.

Pulses are being considered as food of future based on the contemporary lifestyle of sedentary behavior and less physical activity. They have low glycemic index and have been reported to indirectly control the blood glucose level thereby preventing risk of diabetes (Lunde et al. 2011). The indirect control of blood glucose level is achieved through the feeling of satiety (feeling of full stomach), thereby reducing the hunger and increasing the food intake gap (Lunde et al. 2011). The ANFs can be reduced through processing, those are either thermo-labile or water soluble. Soaking the seeds before cooking and draining the water (for water soluble ANFs) or stirring the seeds in dry heat (for thermo-labile ANFs) before cooking makes pulses/ grain legumes ready for consumption.

For the genetic enhancement of chickpea with respect to nutraceutical parameters, target traits should be well defined. Nutraceutical parameters essentially include the nutrition quality and grain quality parameters which determine the consumer preference. Nutrition quality parameter include the concentration of proximate (protein, lipid, carbohydrate) and protective (macro and micronutrients like Fe, Zn, etc.) principles of food. The grain quality traits in chickpea include the grain phenotypical characteristics per se, for example, testa color, cotyledon color, 100 seed weight, seed shape, testa texture, raphae, ribbing, etc. Traits like cooking quality and milling attributes are directly associated with those parameters. The pigments imparting color are essentially antioxidants and their role in health is well documented. Besides, the so-called antinutrition factors with therapeutic properties also come in consideration here. Finally, all those traits have to be combined with high yield and stress (biotic/ abiotic) tolerance so that the concerned nutraceutical enhanced genotype or variety is accepted by the farming community.

5 Importance in Alleviating Chronic Diseases and Malnutrition

Chickpea have been classified as "functional food" that has numerous health benefits like reducing blood glucose, promotes bone health, assist in blood pressure management, promotes cardiac health, helps in reducing cholesterol, promotes mental health, promotes digestive health, and promotes weight management. Chickpea also serves as a cheap source of vitamins, mineral, and numerous bioactive compounds (carotenoids, isoflavones, enzyme inhibitors, etc.) that are essential for healthy living. Chickpeas seeds are rich source of calcium, phosphorus, and iron that

promote bone health and prevents osteoporosis. Chickpea contains choline and selenium two essential nutrients that promote brain and nervous system health.

Chickpea diet has been reported to combat several chronic diseases including type II diabetes, cardiovascular diseases (CVD), and cancers. Chickpea have low glycemic index (GI) and rich in dietary fiber, PUFA, and bioactive compounds (Isoflavones, saponins, and β-Carotene) that lower the risk of CVD and maintains cardiac health. Furthermore, consuming chickpea helps in management of type II diabetes. Its seeds are rich in complex sugar (starch and amylose) that are slowly digested in the small intestine, ensuing low bioavailability of glucose and lowering the demand for insulin. Consumption of complex sugar diet has been proven to enhance glucose tolerance and insulin sensitivity in humans.

Research has indicated that consuming chickpea promote butyrate production. Butyrate belongs to a group of short-chain fatty acids (SCFAs) that has been reported to inhibit cell division and induces apoptosis, thereby preventing development of colorectal cancer. Furthermore, butyrate initiates an irreversible pathway of maturation causing cell death. Lycopene, a bright red colored oxygenated carotenoid is found in chickpea (Jukanti et al. 2012), inhibits development of prostate cancer, lung, and other forms of cancer.

6 Gene Pools

Gene pools of chickpea have been proposed by various authors based on crossability between different species. On the basis of the results derived in wide hybridization, different authors had proposed the gene pool structure of chickpea which are broadly similar. Gene pool structure classifies the annual *Cicer* species and forms the basis on which further wide hybridization breeding program is decided. Harlan and Wet (1971) first proposed the gene pool classes of chickpea which has been validated and contrasted by various authors. Smartt (1984) reviewed the gene pools of legumes and referring to the studies of Ladizinsky and Adler (1976) proposed division of the primary gene pool into GP 1a. and GP 1b., housing *C. arietinum* and *C. reticulatum* respectively. Secondary gene pool (GP 2) was proposed to be comprising of *C. echinospermum*. The tertiary gene pool (GP3) was proposed to be comprising of *C. bijugum, C. pinnatifidum, C. judaicum, C cuneatum, C. yamashitae,* and other annual and perennial species. The gene pool structure proposed by Harlan and Wet (1971) has been validated by Van der Maesen et al. (2007). Here, *C. echinospermum* was placed in primary gene pool. Six annual species have been grouped within the crossable limit. The gene pool structure comprises of *C. arietinum* in the primary gene pool (GP 1a.); *Cicer reticulatum* and *C echinospermum* in GP 1b.; *C. bijugum, C. pinnatifidum,* and *C. judaicum* in GP 2; whereas rest *Cicer* species belong to GP 3. Gene pool structure concluded in the report of by Van der Maesen et al. (2007) is as follows: GP1a: *Cicer arietinum;* GP1b: *Cicer reticulatum* and *C. echinospermum;* GP2: *C. bijugum, C. pinnatifidum, C. judaicum;* GP3: other *Cicer* Species (annual and perennial). An equivalent system was proposed by Redden and Berger (2007), who included *C. arietinum* and *C. reticulatum* in the primary gene pool, *C. echinospermum* in the secondary gene

pool, whereas rest of the annual and perennial species in the tertiary gene pool. Mallikarjuna et al. (2011) modified the above gene pool structure placing *C. arietinum* in primary gene pool (GP 1); *C. reticulatum* and *C. echinospermum* in secondary gene pool (GP 2) and all the rest annual and perennial species in tertiary gene pool (GP 3). Gene pools of *Cicer* proposed by Mallikarjuna et al. (2011) are as following: Primary gene pool: *C arietinum* (Land races and cultivars); Secondary gene pool: *C. reticulatum, C. echinospermum*; Tertiary gene pool: *C. bijugum, C. judaicum, C. pinnatifidum, C. chorassanicum, C. yamashitae, C. cuneatum, C. atlanticum, C. incisum, C. incisum* ssp. serpentinica, C. *floribundum, C. floribundum* var. amanicola*, C. graecum, C. heterophyllum, C. heterophyllum* var. kassianum, *C. uludereensis, C. isauricum, C. montbretii, C. acanthophyllum, C. anatolicum, C. balcaricum, C. baldshuanicum, C. fedtschenkoi, C. flexuosum, C. grande, C. incanum, C. korshinskyi, C. laetum, C. luteum, C. macracanthum, C. microphyllum, C. multijugum, C. nuristanicum, C. paucijugum, C. pungens, C. rassuloviae, C. rechingeri, C. songaricum, C. stapfianum, C. subaphyllum, C. tragacanthoides, C. kermanense, C. mogoltavicum, C. oxyodon, C. spiroceras, C. canariense.*

Out of all the gene pool structures proposed in chickpea, it can be concluded that three annual species, namely, *C. arietinum, C. reticulatum,* and *C. echinospermum* are readily crossable and trait transfer can be attempted through conventional hybridization. In case of wide hybridization involving any other annual or perennial species, interventions like hormonal treatment of the peduncle to prevent abscission, embryo rescue, etc., are essential. Strategic locations often help wide hybridization activities, particularly locations with prolonged congenial temperature window (<25 °C) with normal relative humidity and less wind. It has been observed in the Wide Hybridization Garden Facility at the ICAR-Indian Institute of Pulses Research, Kanpur, that congenial conditions prevailing after pollination enhances the chances for fertilization and further development. Zygotic embryogenesis through embryo culture is a cornerstone of wide hybridization activities. Pod and seed setting in wide hybridization have been long standing issue. Besides, development of the chlorotic mutants like green, partially green, albino, etc., plants from the rescued embryo limits the embryo rescue technique to considerable extent in chickpea. Limited literature and reports are available in embryo rescue in this crop.

7 Diversity Analysis

Jukanti et al. (2012) reviewed the nutritional profile of 100 grams of raw chickpea seeds which approximately provide about 5.0 mg of iron, 4.1 mg of zinc, 138 mg of magnesium, and 160 mg of calcium. Grain protein content of chickpea has been reported to range from 17–22% (Jukanti et al. 2012). Genetic variation for total protein content in grains do exist in the germplasm of chickpea, the maxima being reported >28% in indigenous accessions like T39-1 and T1-A (Singh 1983). Till date, there are limited reports of genetic enhancement of seed protein content in chickpea through conventional breeding or marker assisted interventions. Among

food legumes, protein content has been characterized and genetically enhanced to some extent in soybean, groundnut, pea, etc., (reviewed by Burstin et al. 2011). The amino acid composition of chickpea protein is balanced and is rich in lysine, however concentration of tryptophan and sulphur containing amino acids (methionine, cysteine) is low. The average content of lysine, methionine, and tryptophan in chickpea seed is 7.2 mg, 1.1 mg, and 0.9 mg per gram of protein, respectively. Kaur et al. (2014) reported seed protein content among 30 Indian chickpea cultivars ranged from 16.9% to 29.0% with a mean of 22.6%. Singh (1983), gave a detail account of the work done in ICRISAT and partner institutes between 1976 and 1982 on protein content in chickpea. Protein content of chickpea germplasm panel of 150 lines was estimated following the Kjeldahl method, Technicon auto analyzer method, Dye binding capacity method, and Biuret method over the years and across several locations in India. A comparative study of different methods of protein estimation in chickpea showed equivalent expression across the protocols. Influence of environment on seed protein content was also studied. The lines were found to be stable in protein content over the years. Location had pronounced effect on the protein content of the seeds. Chickpea lines T-1-A, P-3318, and T-39-1 were found to be highest protein containing lines, with a range of 29–32%. The panel had a range of 15–32% in total seed protein content.

Serrano et al. (2017) screened a representative panel of the European chickpea germplasm collection comprising of 79 accessions and observed protein ranging from 14.6% to 23.2%, fibre 3.43–11.11%, fat 3.92–6.98%, and ash ranging from 2.99% to 5.18% respectively. Concentration of resistant starch was comparatively less in the panel (0.13–4.56%). Vitamin E (tochopherol) profile was developed for the germplasm accessions in which α-Tocopherol ranged from 13.42–216.75 mg/g fat, β-Tocopherol was not detected, γ-Tocopherol ranged from 300.86–1078.76 mg/g fat, and δ-Tocopherol ranged from 16.86–57.35 mg/g fat. The concentration of lutein was of a wide range 2.13–28.32 mg g–1 flour. Pigment components like Zeaxanthin also varied significantly, i.e., from 0.13–8.62 mg/g flour and β-Cryptoxanthin was not detected at all. Several accessions were identified with higher concentrations of α-Tocopherol (>200 mg/g). Genotype LEGCA728, with green seed coat and cotyledons, was identified with high lutein concentration (28.32 mg/g). European chickpea accessions were observed to be nutritionally promising and had variability regarding different quality and nutrition parameters, which is the prerequisite of a breeding program. Özer et al. (2010) screened Turkish Kabuli landraces for nutritional quality parameters. Variability pattern observed in the landraces of the place of center of origin of the crop gives an idea of the breeding potential of the traits for which diverse alleles are available. Significant variability was observed for nutrition quality traits like protein (17.55–23.31%), fat (4.45–6.11%), ash (2.54–3.41%), fiber (2.03–4.18%), starch (41.76–49.07%), and 100-seed weight (25.03–51.67 g). Grain quality traits, namely, hydration capacity (0.2585–0.6169 g/ml), swelling index (0.7207–1.1859), swelling capacity (0.15–0.32 ml/seed), cooking time (33–72 min) and seed density (0.8450–1.4800 g/ml) also exhibited considerable variation. Significant correlation was observed between the seed quality and cooking parameters, which has potential for indirect selection for enhancement of cooking quality.

Tripathi et al. (2012) studied the diversity and interrelationships of phenological, physicochemical, and cooking quality parameters in 86 chickpea (*Cicer arietinum* L.) genotypes, including 44 Kabuli and 42 Desi genotypes (varieties and advanced breeding lines). Considerable variation was observed for the grain quality traits, namely, hydration capacity (0.11–0.68 g water/seed), hydration index (0.80–1.21), swelling capacity (0.11–0.7 ml/seed), seed volume (0.1–0.52 ml/seed), and cooking time (38–125 min). Except for hydration index, swelling index and cooking time, Desi and Kabuli types differed significantly in all phonological and quality parameters. Traits like 100-seed weight, hydration capacity, swelling capacity, and seed volume were recorded with high heritability and genetic advance, thereby can be taken as selection parameters. Test weight and seed volume traits were positively correlated with hydration capacity and swelling capacity. This study contrasted the study of Özer et al. (2010) in terms of cooking time trait, for which no correlations were observed, indicating some other factor/loci controlling this trait. This observation may be attributed to the genotype composition of the panel, where on one side Turkish landraces (kabuli type) were studied by Özer et al. (2010) and on the other hand a composite panel of genetically enhanced Desi and Kabuli genotypes were studied. This observation also indicates that cooking properties of Desi and Kabuli grains are different, this may be due to the higher fiber content in the grains of Kabuli genotypes which leads to higher water imbibition, a basic determiner trait for cooking time.

Bhagyawant et al. (2015) carried out biochemical characterization of chickpea accessions for total protein, proline, and antioxidant activity. Significant variation in biochemical composition of chickpea was noted for total protein (100–266 mg/g.), proline (12.2–19.4 mg/100 g), and antioxidant activity. Concentration of Zn and Fe ranged from 0.37 to 0.91 mg/g in the panel. *Cicer reticulatum* accession ILWC-257 was identified with highest Zn (0.66 mg/g) and Fe (2.76 mg/g) and may be used as a potential donor to improve the Fe and Zn content.

8 Mapping of Health-Related Genes

Nutritional quality traits like protein, beta carotene, Fe, and Zn content in seeds of chickpea has been addressed through molecular marker interventions. Molecular marker interventions include association mapping, quantitative trait locus (QTL) mapping, and development of trait specific markers. Abbo et al. (2005) mapped quantitative trait loci (QTLs) for beta carotene and lutein in an interspecific cross with *Cicer reticulatum* accession Cr205 originating from Turkey as donor (0.48 μg/gm). In the inter specific (*Cicer arietinum*/ *Cicer reticulatum*) F_2 population, Abbo et al. (2005) mapped 4 QTLs for beta carotene and single QTL for lutein. The most significant QTL of LOD (Logarithm of odds) >3 was located in an unlinked locus with marker TS 19. Two of three other QTLs reported for beta carotene were located on linkage groups 3B and 1B (B denotes associated linkage group along with the main linkage group 3 and 1). The last QTL for beta carotene was also mapped in an unlinked locus. SSR markers like TA 64, STMS 28, GA11, TA122, and TR 26 were observed to be linked with the beta carotene QTLS and can be utilized for future

marker assisted interventions. Single QTL for lutein concentration was reported on linkage group 8 with LOD 2.4.

Jadhav et al. (2015) carried out association mapping for seed protein content in a varying germplasm panel of 187 chickpea genotypes (comprising of both *Desi* and *Kabuli* types). Nine significant marker-trait associations (MTA) representing four QTLs were identified by the authors considering the entire population. Sub-population analyses identified ten significant MTAs representing five QTLs, however, four of those QTLs were common with the previously identified QTLs in the entire population. Two most significant QTLs were present on LG3 and LG5 and linked with markers TR26 (205 cM) and CaM1068 (195 cM). Twenty-nine candidate genes were identified (gene ontology search) in the region of significant MTAs on LG3. LG3 and LG5 were suggested as possible targets for identification of closely linked markers for protein content in chickpea and for their use in marker assisted genetic enhancement for nutritional quality.

Upadhyaya et al. (2016) carried out genome wide association mapping (GWAS) for Fe and Zn concentration in a diverse panel of 92 chickpea accessions comprising of the Desi and Kabuli genotypes. QTLs for Fe and Zn content in developing seeds were identified in an interspecific mapping population developed from the cross ICC 4958/ICC 8261 (F_7 RIL population with 277 individuals). Eight major genomic regions (for two seasons) governing Fe and Zn concentrations in seed explaining 39.4% combined phenotypic variation on six chromosomes through SNP-based high-resolution QTL mapping in the F_7 RIL population ICC 4958/ICC 826. One QTL was identified each on chromosome 1, 2, , and 7 respectively. Two QTLs were identified on chromosome 3 and 4 respectively. LoD of those QTLs ranged from 6–8.8, identified QTLs explained 16–23% phenotypic variation and the additive component ranged from 4.1 to 8.7. Sixteen genomic loci/genes associated (29% combined phenotypic variance) with seed-Fe and Zn concentrations were identified through GWAS. One genomic region was identified in chromosome 2; two on chromosome 1 and 7; three on chromosome 4 and 5; five genomic regions on chromosome 3. Eleven trait-associated SNPs linked tightly with eight QTLs were validated by QTL mapping and differential expression profiling as well. Rapid integrated genomic strategy led to delineation of novel functional nonsynonymous and regulatory SNP allelic-variants from 16 known/candidate genes. Those included three strong trait-associated genes (encoding late embryogenesis abundant and yellow stripe-like 1 protein and vacuolar protein sorting-associated protein) and eight major QTLs regulating seed-Fe and Zn concentrations.

Ozkuru et al. (2019) reported association mapping for seed Copper, Phosphorus, and Potassium concentrations in a population consisting of 107 *Cicer reticulatum* and 73 *Cicer arietinum* accessions raised at two locations for two seasons. Through genome-based selection, 121,840 SNPs were genotyped across the 180 genotypes. The authors identified eight SNPs which were significantly associated with variations in three nutrient elements (Cu, P and K) in more than two environments.

Wang et al. (2019) mapped QTLs for seed protein content in chickpea in a recombinant inbred line (RILs) population developed from the cross ICC 995 (medium protein)/ICC 5912 (high protein donor). On the basis of multi

environment analysis, one QTL was mapped in each of LG 1 and 6; whereas 2 QTLs were mapped on LG 3, respectively. Significant effect was observed in a QTL identified on LG 3 which explained 52.5% of the phenotypic variances for 100 seed weight, 44.3% for protein content in seeds, and 14.6% for seed shape. Flower color (95.2% phenotypic variance explained) was also observed to be associated with this locus. Alleles of this QTL on LG 3 might be inherited from the high protein donor parent ICC 5912 with blue flower in the RILS which conferred the blue flower color and produced small, round seeds with relatively high protein concentration. The QTL was also observed to be linked with starch accumulation; however, it explained the phenotypic variation for starch content to a lower extent (4%). Protein and starch accumulation being antagonistic to each other, identification of a QTL associated to both protein and starch accumulation is quite intriguing. The study concluded that the genes affecting seed filling at QTL in LG 1 and seed coat development at QTL mapped in LG 3 were responsible for the differences in seed composition and morphology observed in the RIL population.

Fayaz et al. (2022) carried out a GWAS (Genome Wide Association Study) in a diverse core set of 147 chickpea genotypes. The accessions were phenotyped for content of four micro nutrients, namely, Cu, Fe, Mn, and Zn in seeds for two consecutive seasons. The panel was genotyped with 50 K Cicer SNP array. Significant variation was observed for all the parameters in the association panel. During the first season, Zn concentration in seeds ranged 14.56 204.39 ppm (mean 45.31 ppm), Cu concentration ranged 4.23 185.47 ppm (mean 60.99 ppm), Fe concentration ranged from 79.36 to 198.26 ppm (mean 143.24 ppm), and Mn content ranged 9.07 54.29 ppm (mean 25.62 ppm). In the consecutive season, Zn concentration ranged 9.36 205.86 ppm (mean 46.58 ppm), Cu concentration 3.66 190.36 ppm (mean 66.81 ppm), Fe content ranged 100.36 127.36 ppm (mean 132.29 ppm), and Mn content ranged 4.86–122.86 ppm (mean 28.68 ppm) in seeds. However, the authors reported significant differences for all the four micronutrient concentrations ($P \leq 0.05$) across both seasons which posed a challenge in future breeding programs regarding stable donor for respective traits. The mean concentrations of Fe, Zn, Cu, and Mn pooled over two-year data were 146.1 ppm, 45.9 ppm, 63.8 ppm, and 27.0 ppm respectively in seeds. Thirty-five SNPs were identified that are significantly associated with seed Zn, Cu, Fe, and Mn concentration. Five stable MTAs (consistently identified in consecutive seasons) were identified. Six major MTAs identified which explained more than 15% phenotypic variation of respective trait and three MTAs, both major and stable in nature were reported.

9 Breeding Strategies and Genetics

Smartt et al. (1975) outlined the thrust areas for a breeding program targeting for enhancement of seed protein content. The three thrust areas included improvement of the overall protein content, improvement of the amino acid profile of the seed storage proteins, and improvement of the digestibility. Attempts to mobilize the genes for high protein content from the improved landrace background of T39-1

(ICC 5912) or T1-A to the high yielding backgrounds of chickpea by hybridization and consequent handling of the segregating generations are currently in progress. Smartt et al. (1975) also outlined the selection approaches for enhancing protein parameters, those were, selection for a specific desirable amino acid profile, selection for total protein content, and negative selection against any particular amino acid profile. Selection for quality parameters needs handling of large populations and developing the SDS-PAGE profiles of seed tissue of breeding lines of a large population is a cumbersome process. Vijayalakshmi et al. (2001) carried out primary work to study the inheritance pattern of the seed protein content trait in chickpea. T39-1, a high protein containing line was crossed with moderate and low protein parental lines P9623 and RS11. Low protein content was been reported to be dominant over high protein content trait. Oligogenic inheritance was observed in F_2 and transgressive segregant were also observed in advanced generations.

The genotype × environment interaction studies with protein content in chickpea are interesting and provide various ways to conceptualize the approaches for genetic enhancement of the trait. Singh et al. (1993) studied the genotype × environment interaction of the trait protein content in chickpea seeds in five set of trials evaluated in three locations for two winter seasons of the Mediterranean region. Effects of genotype × location and genotype × season were observed to be small on the seed protein content indicating genetic component for the trait. The genotypes were observed to be stable across the environment and season for protein content (season to season variation was 0.1–1%) and had negligible effect of sowing date in the Mediterranean region.

In another report, the performance of chickpea genotypes for protein content at eleven locations across India for two seasons was studied Singh et al. (1983). Pronounced effect of location was observed for variance of protein content, whereas effect due to genotypes on the variance was low. Insignificant effect of cultivars × location on variance of protein content was observed suggesting breeding potential of the seed protein content trait in chickpea. The study suggested breeding material adapted to the location for genetic enhancement for protein content in chickpea. Frimpong et al. (2009) reported significant genotype × environment interaction effects on starch, amylose, and protein concentrations in chickpea. The correlations identified in this study were interesting, in both *Desi* and *Kabuli* genotypes, starch concentration was negatively correlated with the protein content in the seed (r Kabuli $= -0.25$, $P < 0.05$; r Desi $= -0.16$, $P < 0.05$). Though not universal, but this finding gives a direction to understand the protein accumulation physiology in chickpea. The authors reported negative correlation of yield and protein content. The chickpea landrace with high protein T39-1 (ICC 5912) which was identified by Singh (1983) had considerable heat susceptibility. It is evident from those findings that linkage drag may exist while selecting for protein content in chickpea. Such donors need to be crossed with diverse recipient parents (check varieties, high in yield and biotic/ abiotic stress resistance/tolerance), and large F_2 populations need to be grown to address the linkage drags and identify the desirable recombinants.

Wang et al. (2017) reported genotype × environment interaction for different grain quality parameters like seed weight, total starch, amylose, and crude protein in

237 chickpea genotypes (180 *Desi*, 49 *Kabuli,* and 8 *pea* shaped) for two seasons grown under field conditions and greenhouse. Kabuli genotypes showed higher total starch (34.1–63.2%) and crude protein (15.7–28.0%) concentrations; whereas, amylose concentration (27.1–41.6% of total starch) was higher in *Desi* genotypes. Intriguing inter trait associations were observed in this diverse panel, on which correlated selection strategies can be worked out. Significant negative association of seed weight with amylose and crude protein was observed. Seed weight was positively associated with total starch. Amylose content was significantly negatively correlated to starch. Crude protein was positively correlated to amylose content and negatively correlated with starch content. Broad sense heritability (H^2) was observed for seed weight (0.70), whereas low to moderate H^2 was observed for starch (0.12–0.51), crude protein (0.15–0.37), and amylose (0.10–0.14) content. This finding is very important and a major factor for working out selection indices for genetic enhancement of protein content. Based on accumulation of amylose (early podding-seed setting stage) and its conversion to starch (late podding – grain filling stage) phenological traits may be targeted for association and indirect selection for high seed protein content. However, this finding needs to be tested in different germplasm panels and to be tested in diverse locations for their universality because such findings are often result off type genotypes.

Wang et al. (2017) validated the findings of Frimpong et al. (2009) and recommended chickpea genotypes with high crude protein and amylose content can be utilized in breeding programs. Cobos et al. (2016) studied genotype × environment interaction in chickpea genotypes grown in winter and spring sowings in four different locations in the Mediterranean region and reported high oil, acid detergent fiber (ADF) but low protein with respect to the spring harvest. The authors reported pronounced genotype effect on oil, acid detergent fiber (ADF), and protein content. Environmental effect was more in amylose and amylopectin concentration. Cobos et al. (2016) advocated for indirect selection for the grain nutritional parameters and suggested selection of lines on the basis of sensory/ nutritional/ grain quality traits in advanced generations only. No approach to handle the segregating populations in early generation has been reported in the limited literature surveyed, which authors consider a major research gap. No multitrait selection model was identified in the literature for genetic enhancement of nutritional and quality parameters in pulse crops where segregation in seed traits up to late generation poses a problem for varietal release.

Gaur et al. (2016) reported multiple gene control of protein content and its association with other traits. Contrasting cross was developed utilizing high protein donor ICC 5912 (29.2% protein) which had small seed, grey seed coat, and blue flower with a *Kabuli* line ICC 17109 (20.5% protein) having contrasting traits. Seed protein content and other contrasting traits were controlled by multiple genes and had normal F_2 distribution. Flower color and testa color traits showed recessive epistasis (9:3:4) in F_2. Pleiotropy was reported for flower color and testa color traits as well. High protein trait was associated with blue flower trait in the segregating lines of F_2 population. Authors reported negative correlation with seed size and yield. This may be attributed to the heat susceptibility of ICC 5912. Notably, ICC

8397 (T1-A, landrace), ICC 5912 (T39-1, landrace) have been confirmed for high protein content in seeds (~29%, Kjeldahl's method), and ILWC 237 (wild germplasm accession, *Cicer reticulatum*) has been confirmed for high Fe, Zn concentration. Those donors are being utilized in crossing program for generation of breeding materials.

10 Next-Generation Breeding

Around 3 billion people in world suffer from malnutrition caused due to nutritional deficiencies of iron or zinc. Chickpea being rich in iron (Fe), zinc (Zn), and selenium (Se) can be an alternative diet supplement in order to counter malnutrition. Since last decades several attempts have been made to counter "hidden hunger" (deficiency of mineral nutrients and micro nutrients) and biofortified chickpea varieties have been developed expressing higher levels of these minerals (Vandemark et al. 2018). Developing biofortified cultivars that stably over-express selected minerals requires genetic information, environmental interaction, and its effect on mineral concentrations (Ray et al. 2014). Biofortification of chickpea employing next-generation breeding to enhance nutritional value is vital for combating these issues associated with global hunger and malnutrition. The introduction of new breeding technologies is expected to contribute significant improvements in crop production. Although, conventional breeding methods have developed more than 200 improved varieties of chickpeas, there is still scope for better productivity. Conventional breeding is time consuming and laborious process that requires substantial amount of space for crossing of desirable crops. The long duration of cropping season and seed to seed cycle are the major limitation in the conventional breeding method. Integrated uses of modern genomic resources and speed breeding methods can produce the desired varieties efficiently. With the advancement of modern genomic tools and analysis of high-resolution phenotypic and genetic data has assisted in trait associated gene identification. Genomic tools have been developed for marker identification employed in marker-assisted back-crossing for the production of improved chickpea varieties.

QTLs associated with carotenoid components (violaxanthin, lutein, zeaxanthin, β-cryptoxanthin, and β-carotene) in chickpea was derived from crosses between cultivars with green and yellow cotyledon colors using high-performance liquid chromatography (HPLC). The result indicated a total of 1068 bin markers that were derived from the 50 K Axiom CicerSNP array (mapped onto eight linkage groups (LGs)). A total of eight QTLs in the "CDC Jade" × "CDC Frontier" population, including two each for β-carotene and zeaxanthin and single QTLs for total carotenoids, β-carotene, violaxanthin, and β-cryptoxanthin were identified. Similarly, 694 bin markers were mapped onto eight LGs and one partial LG in the "CDC Cory" × "CDC Jade" population. Various carotenoids including β-cryptoxanthin, β-carotene, violaxanthin, lutein, and total carotenoids were identified on linkage group 8. In another "ICC4475" × "CDC Jade" derived population indicated presence of 581 bin markers in the third population. Total five QTL one for β-carotene

and four others, one each of β-carotene, lutein, β-cryptoxanthin, and total carotenoids, were identified in this population. Interestingly, highest phenotypic variation was observed in the β-carotene QTLs, ranging from 58% to 70% in all the three crossed populations.

The total isoflavone content (TIC) in chickpea is reported to vary 153 to 340 mg/100 g of chickpea (Singh et al. 2017). Formononetin and biochanin A are the major isoflavones present in chickpea while biochanin A-7-O-β-D-glucoside, genistein, calycosin, trifolirhizin, sissotrin, and ononin are present in smaller proportion (Kashiwagi et al. 2015). Isoflavones promotes uterine growth and prevent bone loss. Furthermore, isoflavones induces estrogen responsive element (ERE)-promoter that promotes estrogenic activity in cells.

Chickpea has numerous nutritional and health benefits along with some associated processing issues, like the presence of antinutrients, prolonged cooking time, and poor digestibility. These antinutritional elements include tannin, phytic acid, trypsin inhibitor, chymotrypsin inhibitors, and alpha amylase inhibitors that inhibit the nutritional availability. Bioavailability of the nutrient present in chickpea is the function of the interaction between antinutrients and nutrients. These antinutritional factors usually form complexes with the nutritional elements and inhibit their absorption. Among the various antinutritional factors found in chickpea, tannins and phytate are the most prevalent. Tannins are polyphenols that have been reported to interact with proteins and enzymes to form tannin-protein complexes, that severely diminishes nutritional bioavailability. Phytate is an inositol hexa-phosphate that interacts with minerals resulting in their reduced absorption. Minerals like, zinc, calcium, magnesium, and iron found in chickpea are present in traces, and the phytic acid found in chickpea interact with these minerals and inhibits their absorption. Similarly, trypsin and chymotrypsin inhibitors interact with hydrolytic enzymes and obstruct their activity. The alpha-amylase inhibitors impede the activity of amylase enzyme and obstruct digestion and absorption of starch. Hence, it is essential to reduce the antinutritional factors for biofortification of chickpeas. Major biofortification efforts are significantly focused on micronutrients, however these techniques can be applied to other nutritional traits Roorkiwal et al. (2021). Similarly, Chickpea is rich in linoleic acid (LA; ω-6), compared to the α-linolenic acid (ALA; ω-3), the other essential fatty acid. Thus, breeding efforts to enrich these essential fatty acids is important aspect; however, these traits are governed by multiple genes families making the process complicated.

11 Biotechnological Interventions

Genetic engineering (GE) approach for biofortification with Iron (Fe) in chickpea employing combination of the chickpea nicotianamine synthase 2 (*CaNAS2*) and soybean ferritin (*GmFER*) genes was reported (Tan et al. 2018). The *CaNAS2* gene is responsible for Fe transportation while the soybean derived gene *FER* help in storage of iron. The experiment was conducted using three commercial Kabuli cultivars (Genesis090™, Kalkee™, and PBA Monarch) and three commercial Desi cultivars

(PBA Boundary, CICA0912, and PBA HatTrick). Similar Fe levels were reported among all the tested cultivars and lower Fe levels were obtained where soil Fe bioavailability was low. Analysis of transformants indicated doubled levels of nicotianamine (NA) compared to the control while the Fe levels remained unchanged. Higher NA level is linked with high Fe bioavailability that can play an active role in overcoming the issues associated with the bioavailability of Fe in the presence of inhibitors found in chickpea; however, further study was required for confirmation. In another report, transgenic chickpea expressing CaWRKY31: CaCKKX6 were found to have increased root mass, resulting in increased nutrient content in seeds of the transgenic lines: Zn (27–62%), Cu (26–61%), and Fe (22–48%) (Khandal et al. 2020).

Genome editing by targeting gene for cytokinin dehydrogenase was also suggested for biofortification (Mahto et al. 2022), owing to its linkage with stress regulation (Cortleven et al. 2019) and plant mineral concentration (Gao et al. 2019).

12 Conclusion

Chickpeas are very important for nutritional security, and its characterization and enrichment are immediate priority. However, accomplishments in quality breeding in chickpea are limited and yet to head to a product delivery phase. Nutritional profiles based on data generated for protein, fats/oils, carbohydrate, fibers, folate, beta carotene, and different elements in different chickpea genotypes are available in the literature. Majority of such reports have taken very limited number of genotypes and has worked on the quality of the protein (amino acid profile). The methods adopted in those reports are the conventional biochemical protocols. Few studies on diversity available in the germplasm for quality parameters like protein, iron, and zinc content are available in pulse crops which will be the foundation for future plant breeding. Research status of the protein aspect in chickpea is largely confined to the amino acid profiling or protein fraction characterization in the germplasms. Nutrition quality breeding is a team work of breeders, biochemists, physiologists, engineers, and food technologists to scale up the present research to product delivery.

References

Abbo S, Molina C, Jungmann R, Grusak MA, Berkovitch Z, Reifen R, Kahl G, Winter P, Reifen R (2005) Quantitative trait loci governing carotenoid concentration and weight in seeds of chickpea (*Cicer arietinum* L.). Theor Appl Genet 111(2):185–195

Bhagyawant SS, Gupta N, Shrivastava N (2015) Biochemical analysis of chickpea accessions Vis-a-Vis; zinc, iron, total protein, proline and antioxidant activity. Am J Food Sci Technol 3 (6):158–162

Burstin J, Gallardo K, Mir RR, Varshney RK, Duc G (2011) Improving protein content and nutrition quality. In: Pratap A, Kumar J (eds) Biology and breeding of food legumes. CAB International, Wallingford, pp 314–328

Cobos MJ, Izquierdo I, Sanz MA, Tomás A, Gil J, Flores F, Rubio J (2016) Genotype and environment effects on sensory, nutritional, and physical traits in chickpea (*Cicer arietinum* L.). Span J Agric Res 14(4):0709

Cortleven A, Leuendorf JE, Frank M, Pezzetta D, Bolt S, Schmülling T (2019) Cytokinin action in response to abiotic and biotic stresses in plants. Plant Cell Environ 42:998–1018

Van der Maesen LJG, Maxted N, Javadi F, Coles S, Davies AMR (2007) Taxonomy of the genus *Cicer* revisited. In: Yadav SS, Redden R, Chen W, Sharma B (eds) Chickpea breeding and management. CABI, Oxfordshire, pp 14–46

Duke JA (1981) Handbook of legumes of world economic importance. Plenum Press, New York, pp 52–57

FAOSTAT (2020). https://www.fao.org/faostat/

Fayaz H, Tyagi S, Wani AA, Pandey R, Aktar S, Bhat MA, Chitkineni A, Varshney RK, Thudi M, Kumar U, Mir RR (2022) Genome-wide association analysis to delineate high-quality SNPs for seed micronutrient density in chickpea (*Cicer arietinum* L.). Sci Rep 12:11357

Frimpong A, Sinha A, Tar'an B, Warkentin TD, Gossen BD, Chibbar RN (2009) Genotype and growing environment influence chickpea (*Cicer arietinum* L.) seed composition. J Sci Food Agric 89(12):2052–2063

Gao S, Xiao Y, Xu F, Gao X, Cao S, Zhang F et al (2019) Cytokinin-dependent regulatory module underlies the maintenance of zinc nutrition in Rice. New Phytol 224(1):202–215

Gaur PM, Singh MK, Samineni S, Sajja SB, Jukanti AK, Kamatam S, Varshney RK (2016) Inheritance of protein content and its relationships with seed size, grain yield and other traits in chickpea. Euphytica 209:253–260

Harlan JR, Wet JMJ (1971) Toward a rational classification of cultivated plants. Taxon 20(4):509–517

Jadhav AA, Rayate SJ, Mhase LB, Thudi M, Chitikineni A, Harer PN, Jadhav AS, Varshney RK, Kulwal PL (2015) Marker-trait association study for protein content in chickpea (*Cicer arietinum* L.). J Genet 94:279–286

Joshi PK, Parthasarathy Rao P (2016) Global and regional pulse economies current trends and outlook. Washington, D.C, IFPRI

Jukanti AK, Gaur PM, Gowda CL, Chibbar RN (2012) Nutritional quality and health benefits of chickpea (*Cicer arietinum* L.): a review. Brit J Nutr 108:S11–S26

Kashiwagi J, Krishnamurthy L, Purushothaman R, Upadhyaya HD, Gaur PM, Gowda CLL, Ito O, Varshney RK (2015) Scope for improvement of yield under drought through the root traits in chickpea (*Cicer arietinum* L.). Field Crop Res 170:47–54

Kaur S, Kaur S, Gupta AK, Kaur N, Javed M (2014) Biochemical and nutritional characterization of chickpea (*Cicer arietinum* L.) cultivars. Indian J Agric Sci 84(4):479–486

Khandal H, Gupta SK, Dwivedi V, Mandal D, Sharma NK et al (2020) Root specific expression of chickpea cytokinin oxidase/dehydrogenase 6 leads to enhanced root growth, drought tolerance and yield without compromising nodulation. Plant Biotechnol J 18(11):2225–2240

Ladizinsky G (1975) A new *Cicer* from Turkey. Notes of the Royal Botanic Garden Edinburgh 34: 201–202

Ladizinsky G, Adler A (1976) The origin of chickpea *Cicer arientinum* L. Euphytica 25:211–217

Lunde MSH, Hjellset VT, Holmboe-Ottesen G, Høstmark AT (2011) Variations in postprandial blood glucose responses and satiety after intake of three types of bread. J Nutr Metabol 2011: 437587

Mahto RK, Ambika Singh C, Chandana BS, Singh RK, Verma S et al (2022) Chickpea biofortification for cytokinin dehydrogenase via genome editing to enhance abiotic-biotic stress tolerance and food security. Front Genet Sec Plant Genom 13:900324

Mallikarjuna N, Coyne C, Cho S, Rynearson S, Rajesh PN, Jadhav DR, Muehlbauer FJ (2011) Chickpea. In: Kole C (ed) Wild crop relatives: genomic and breeding resources. Springer, Berlin/Heidelberg, pp 63–82

Mittal R, Nagi HPS, Sharma P, Sharma S (2012) Effect of processing on chemical composition and antinutritional factors in chickpea flour. J Food Sci Eng 2:180–186

NNMB (2006) Diet and nutritional status of population and prevalence of hypertension among adults in rural areas. National Nutrition Monitoring Bureau Technical Report No: 24. National Institute of Nutrition, Hyderabad. https://www.nin.res.in/downloads/NNMBReport06Nov20.pdf. Accessed on 12.01.2020

Özer S, Karaköy T, Toklu F et al (2010) Nutritional and physicochemical variation in Turkish kabuli chickpea (*Cicer arietinum* L.) landraces. Euphytica 175:237–249

Ozkuru E, Ates D, Nemli S, Erdogmus S, Karaca N et al (2019) Association mapping of loci linked to copper, phosphorus, and potassium concentrations in the seeds of *C. arietinum* and *C. reticulatum*. Genomics 111(6):1873–1881

Ray H, Bett K, Taran B, Vandenberg A, Thavarajah D, Warkentin T (2014) Mineral micronutrient content of cultivars of field pea, chickpea, common bean, and lentil grown in Saskatchewan, Canada. Crop Sci 54:1698–1708

Redden R, Berger J (2007) History and origin of chickpea. In: Yadav S et al (eds) Chickpea breeding and management. CABI, Wallingford, pp 1–13

Roorkiwal M, Pandey S, Thavarajah D, Hemalatha R, Varshney RK (2021) molecular mechanisms and biochemical pathways for micronutrient acquisition and storage in legumes to support biofortification for nutritional security. Front Plant Sci 1210.3389/fpls.2021.682842

Serrano C, Carbas B, Castanho A, Soares A, Patto MCV, Brites C (2017) Characterisation of nutritional quality traits of a chickpea (Cicer arietinum) germplasm collection exploited in chickpea breeding in Europe. Crop Pasture Sci 68(11):1031–1040. https://doi.org/10.1071/CP17129

Singh U (1983) Progress Report 1, grain quality and biochemistry: chickpea and pigeonpea protein content. International Crop Research Institute for Semi-Arid and Tropics, Hyderabad

Singh U, Kumar J and Gowda CLL (1983) The protein content of chickpea (Cicer arietinum L.) grown at different locations. Qual Plant Plant Fd Hum Nut 32:179–184. https://doi.org/10.1007/BF01091338

Singh U (1988) Antinutritional factors of chickpea and pigeonpea and their removal by processing. Plant Foods Hum Nutr 38:251–261

Singh KB, Bejiga G, Malhotra RS (1993) Genotype-environment interactions for protein content in chickpea. J Sci Food Agric 63:87–90

Singh KB, Pundir RPS, Robertson LD, van Rheene HA, Singh U, Kelley TJ, Rao PP, Johansen C, Saxena NP (1997) Chickpea. In: Fuccillo D, Sears L, Stapleton P (eds) Biodiversity in trust. Cambridge University Press, pp 100–113

Singh B, Singh JP, Singh N, Kaurb A (2017) Saponins in pulses and their health promoting activities: a review. Food Chem 233:540–549

Smartt J (1984) Gene pools in grain legumes. Econ Bot 38:24–35

Smartt J, Winfield PJ, Williams D (1975) A strategy for the improvement of protein quality in pulses by breeding. Euphytica 24:447–451

Tan GZH, Das Bhowmik SS, Hoang TML, Karbaschi MR, Long H, Cheng A et al (2018) Investigation of baseline iron levels in Australian chickpea and evaluation of a transgenic biofortification approach. Front Plant Sci 9:788

Toker C, Berger J, Eker T, Sari D, Sari H, Gokturk RS, Kahraman A, Aydin B, von Wettberg EJ (2021) Cicer turcicum: a new cicer species and its potential to improve chickpea. Front Plant Sci 12:662891

Tripathi S, Sridhar V, Jukanti AK, Suresh K, Rao BV, Gowda CLL, Gaur PM (2012) Genetic variability and interrelationships of phenological, physicochemical and cooking quality traits in chickpea. Plant Genet Resour Characteriz Utiliz 10(3):194–201

U.S. Food and Drug Administration (2021) Calories count. Report of the Working Group on Obesity

Upadhyaya HD, Bajaj D, Das S, Kumar V, Gowda CLL, Sharma S, Tyagi AK, Parida SK (2016) Genetic dissection of seed-iron and zinc concentrations in chickpea. Sci Rep 6:24050

Van der Maesen LJG (1972) Cicer L. a monograph of the genus, with special reference to the chickpea (*Cicer arietinum* L.), its ecology and cultivation, vol 72. Mededlingen landbouwhogeschool (Communication Agricultural University), Wageningen, pp 9–136

Van der Maesen LJG (1987) Origin, history and taxonomy of chickpea. In: Saxena MC, Singh KB (eds) The chickpea. CAB International, Cambridge, pp 11–34

Vandemark GJ, Grusak MA, McGee RJ (2018) Mineral concentrations of chickpea and lentil cultivars and breeding lines grown in the U.S. Pacific Northwest. Crop J 6:253–262

Vavilov NI (1951) The origin, variation immunity and breeding of cultivated plants. Chron Bot 13-1 (6):26–38

Via VD, Zanetti ME, Blanco F (2016) How legumes recognize rhizobia. Plant Signal Behav 11: 1120396

Vijayalakshmi NVS, Kumar J, Rao TN (2001) Inheritance of protein content in chickpea. Legum Res 24(1):28–31

Wang R, Gangola MP, Jaiswal S, Gaur PM, Bågaa M, Chibbar RN (2017) Genotype, environment and their interaction influence seed quality traits in chickpea (*Cicer arietinum* L.). J Food Compos Anal 63:21–27

Wang R, Gangola MP, Irvine CM, Gaur PM, Båga M, Chibbar RN (2019) Co-localization of genomic regions associated with seed morphology and composition in a desi chickpea (Cicer arietinum L.) population varying in seed protein concentration. Theor Appl Genet 132(4):1263–1281. https://doi.org/10.1007/s00122-019-03277-5

Wang J, Yu H, Yili A, Gao Y, Hao L, Aisa HA, Liu S (2020) Identification of hub genes and potential molecular mechanisms of chickpea isoflavones on MCF-7 breast cancer cells by integrated bioinformatics analysis. Ann Transl Med 8(4):86

WHO (2007) Proteins and amino acids in human nutrition. WHO/FAO/UNU expert consultation. Technical Report Serial No. 935. World Health Organization. WHO, Geneva

Nutrient-Dense Pea (*Pisum sativum* L.): Genetics and Genomics-Mediated Developments

A. K. Parihar, G. P. Dixit, Amrit Lamichaney, Arpita Das, Kuldeep Tripathi, Neetu Singh, Aravind Konda, DebJyoti Sen Gupta, Surendra Barpete, Sanjeev Gupta, and Abhimanyu Sarker

Contents

A. K. Parihar (✉) · G. P. Dixit · A. Lamichaney · N. Singh · A. Konda · D. S. Gupta
ICAR-Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India
e-mail: gpdixit1966@gmail.com; amrit.lamichaney@icar.gov.in; neeturajawat@gmail.com; aravindbio@gmail.com; debgpb@gmail.com

A. Das
Department of Genetics & Plant Breeding, Bidhan Chandra Krishi Viswavidyalaya, Mohanpur, Nadia, West Bengal, India
e-mail: arpitacoh@gmail.com

K. Tripathi
Division of Germplasm Evaluation, ICAR-National Bureau of Plant Genetic Resources, New Delhi, India
e-mail: Kuldeep.Tripathi@icar.gov.in

S. Barpete
Scientist, Food Legumes Research Platform (FLRP), International Centre for Agricultural Research in the Dry Areas (ICARDA), Sehore, Madhya Pradesh, India
e-mail: surendrabarpete@gmail.com

S. Gupta
Assistant Director General, Oil seeds & Pulses, Indian Council of Agricultural Research, New Delhi, India
e-mail: saniipr@rediffmail.com

A. Sarker
John Innes Centre, Norwich, UK
e-mail: abhimanyu.sarkar@niab.com

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_25

Abstract

Being nutritionally dense legume crop, pea holds great promise toward nutritional security of resource poor folks. Among the legume crops, pea is well recognized to be rich in easily digestible protein coupled with an admirable amount of macro- and micronutrients. It offers a wide range of health advantages due to excellent quality of starch, dietary fiber and reasonable source of flavonoids and phenolic compounds. In this chapter, authors deliberated about the nutritional configuration of peas and its nutraceutical enhancement accomplished through various agronomic and genetic biofortification strategies. In addition the chapter also elaborated about the arsenal of genetic resources; trait specific genotypes for nutritional traits and their genetics mechanism; status of markers and molecular mapping of health-related (HR) genes and QTLs; map-based cloning of HR genes/QTLs; omics approaches in relation to HR genes: genomics-aided breeding for HR traits; genome editing; gene stacking; TILLING; future strategies for modulating HR genes in pea: current status and future possibilities for genetic engineering application for HR traits and role of bioinformatics. Overall very limited efforts so far have been made toward the exploitation of existing variability for nutritional traits. Hence, there is an urgent need of large-scale high-throughput phenotyping to unveil the existing genetic variability and their subsequent utilization in regular breeding program to develop nutrient-enriched pea. However, in the case of pea, genomic resources have not been well developed and exploited for nutritional traits, therefore, recent omics-based approaches need to be embraced to facilitate the identification of genes/QTLs to develop the robust markers' platform to accelerate marker-assisted breeding program. Fortunately, the recently accomplished pea reference genome assembly offered several contigs, transcripts, markers, and GBS platforms for their further annotation. The transcriptomics, proteomics, and metabolomics analysis have not been used considerably for nutrition-related genes. However, few valuable transcripts' atlas has been generated recently that would further boost reverse genetic studies, fine mapping, allele mining, and identification of candidate genes. To further accelerate nutraceutical improvement in pea induced mutagenesis; genome

editing; gene silencing, bioinformatics, and transgenic approaches need to be adopted in regular breeding programs. The advent of affordable genomic tools and their access will speed up the creation of novel pea cultivars with refined HR genes for the benefit of resource poor mankind.

Keywords

Nutraceutical · Biofortification · Conventional approaches · Genetic resources · Genomics · Genome editing · Transgenic

1 Introduction

Pea (*Pisum sativum* L.) is a nutritionally dense annual herbaceous post rainy season legume crop grown around the world for food, fodder, and feed (Parihar et al. 2020, 2022c; Lamichaney et al. 2021a, b). Fresh green seeds and pods are consumed as vegetables whereas mature dry seeds are used in diverse forms, i.e., whole, split, and flour (Parihar et al. 2023). Moreover, it is considered an integral part of plenty of cuisine and confectioneries, for example, *chat, soup, dal, chhola,* vegetables, *snacks, stew,* and flour (Parihar et al. 2014a, b). Being an admirable source of protein, starch and fibre, it is extensively used as an important constituent in several food products-based industries (Dixit et al. 2014; Gupta and Parihar 2015). Pea seeds are nutritionally enriched and have 15–35% protein along with essential amino acids like lysine and tryptophan. Being a legume crop it is an important component of crop rotation as it disturbs the disease and pest life cycle, offers nitrogen, proliferate the level and performance of soil microorganism, enhances soil aggregation and improves soil water holding capacity. Pea is the fourth most important legume crop at global level occupying around 9.80 Mha areas with a production of about 34 Mt. The recent trend of area and production worldwide is illustrated in Fig. 1.

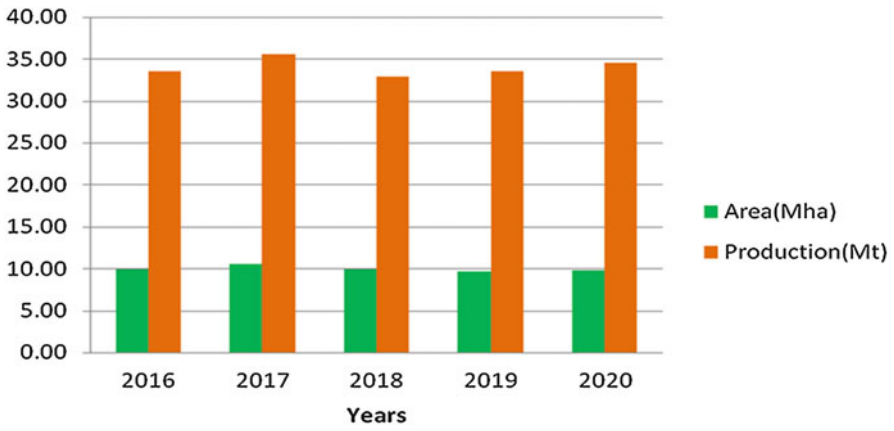


Fig. 1 Recent trends of area and production of peas at global level

In general, food legumes are enriched with various nutrients and deliver a greater amount of essential nutrients relative to cereals and tuber crops (Harmankaya et al. 2010; Blair et al. 2011). Among the legume crops, pea seeds have been well recognized to be rich in easily digestible protein coupled with an admirable amount of macro and micronutrients. Approximately 21–32% of dry seeds are composed of proteins that predominantly contained storage proteins, i.e., globulin and albumin (Tzitzikas et al. 2006). Besides protein, it is a decent source of essential amino acids like methionine, arginine, and valine. It is an excellent source of carbohydrates which ranges from 56% to 74% and is rich in amylase content which leads to slow digestion of starch. The seed coat and cotyledon are rich in dietary fibre, i.e., water unsolvable and solvable fibre (Tosh and Yada 2010). The most dominating macronutrient in pea is potassium followed by phosphorus, magnesium, and calcium. It has also been acknowledged as an excellent source of other micro-nutrients for instance iron, zinc, selenium, and molybdenum. Furthermore, it has an appreciable amount of vitamin B9 (folate), vitamin C (Ascorbic acid), vitamin B1 (thiamin), B6 (pyridoxine), B3 (Niacin), and B2 (riboflavin). Corresponding to other legumes, it is also containing several phytochemicals such as carotenoids, chlorophyll, flavonoids, saponins, and oxalates (Parihar et al. 2016).

Malnutrition has become a severe problem owing to the insufficient supply of nutritionally balanced diet to the resource-poor folks of developing countries. Nutritionally, the pea is a precious food, which can cater the nutritious dietary requirements of about 80–90 crores of malnourished folks globally. In the case of developing countries, approximately 2 billion people are facing malnourishment, whereas above 3 billion folks are deprived of essential micronutrients worldwide. Micronutrient scarcity causes numerous health-related issues, viz., anemia, intellectual debilities, poor immunity, respiratory diseases, and recedes work efficiency. In general, rural habitats of developing nations are mainly having shortages of iron, zinc, selenium and iodine, and globally around 60, 33, and 15% of inhabitants are malnourished for Fe, Zn, and Se, respectively (Hotz and Brown 2004). Besides, vitamin A, Zn, Fe, and/or I deficiencies all together cause approximately 20% mortality in kids of less than 5 years. Deficiency for these micronutrients mostly exists in the folks who are dependent on plant-based diets to cater to their nutritional needs in which bioavailability of iron is less as compared to nonvegetarian diets. Fortunately, the pea is enriched with iron and zinc and therefore, could tackle the prevailing deficiencies for these micronutrients (Demirbas 2018).

This crop offers a wide range of health advantages owing to its excellent quality of starch and dietary fibre which makes it a low glycemic index (GI) food and plays a crucial role in type 2 diabetes inhibition and administration. Besides, consumption of intact peas, flour, and its byproduct checks the arbitrary mounting of blood sugar level. The solvable dietary fibre, niacin, and saponin compounds also reduce the threat of cardiovascular disease by plummeting and stabilizing the blood cholesterol levels (Ekvall et al. 2006). Interestingly, flavonoids also play an instrumental role in heart disease management by inhibiting blood platelet aggregation and providing antioxidant shelter to bad cholesterol. Being a reasonable source of flavonoids and phenolic compounds it has cancer combating and antioxidant properties. The insoluble fibers of peas also improve bowel movement which is vital in the avoidance of

constipation and other gastrointestinal diseases including bowel cancer. The raffinose and other galactose-encompassing oligosaccharides may play the role of prebiotic in the intestine which improves the overall gastrointestinal functions (Fernando et al. 2010). The aforesaid attributes collectively designated pea as an important food article, which can fulfill the dietary needs of people of developing and underdeveloped nations.

2 Nutritional Profile of Pea

The nutritional configuration of peas proved it as an admirable source of carbohydrates, proteins, macro and micronutrients, vitamins, and essential amino acids (Parihar et al. 2022a, b). The existing variability for various nutritional elements is presented in Table 1. Cotyledon is the reservoir of nutrients in pea, while embryo and seed coat contributed to less than 10% of the total nutritive value. The carbohydrates are an integral part of pea seeds and are present in sizable amounts accounting for about 56–74% of the dry matter which is made up of oligosaccharides, monosaccharides, polysaccharides, and disaccharides (Dahl et al. 2012). The starch content accounted for 27.6–57.23% of the pea seed dry matter (Tzitzikas et al. 2006). Starch is grouped into three classes, i.e., slowly digestible starch (SDS), rapidly digestible starch (RDS), and resistant starch (RS) considering the digestion ability of glucose and its absorption efficiency into the gastrointestinal territory (Singh et al. 2017). Starch is made up of amylose and amylopectin and their proportion plays an important character in starch digestibility and after-meal glucose response. In general, pea contains 20.7–38.0% amylase which reduces digestion ability with the release of glucose making it a low GI food (Dahl et al. 2012). Pea comprises of extensive quantity of RFOs and other oligosaccharides, of them, stachyose, total a-D-galactosides, verbascose, raffinose, and sucrose content oscillated from

Table 1 Genetic variability for nutritionally important parameters in peas

S. No	Component	Range	References
1	Protein (%)	13.7–38.3	Tzitzikas et al. (2006); Pandey and Gritton (1975)
2	Starch (%)	20–71	Guillon and Champ (2002); Bhattacharyya et al. (1990)
3	Amylose (%)	20.7–90	Dahl et al. (2012)
4	TDF (%)	10.4–28.0	Li et al. (2002); Chen et al. (2013)
5	SDF (%)	2.0–19.8	Tosh and Yada (2010); Chen et al. (2013)
6	IDF (%)	3.3–15.0	Chen et al. (2013); Tosh and Yada (2010)
7	TOS (mg g^{-1})	37.7–177.6	Gawłowska et al. (2017)
8	Iron (mg kg^{-1})	21.90–320.9	Harmankaya et al. (2010); Demirbas (2018)
9	Zinc (mg kg^{-1})	11.3–	Demirbas (2018); Grusak and Cakmak (2005)
10	Folate (mg kg^{-1})	0.1–0.7	Zhang et al. (2018); Han and Tyler (2003)

SDF: Soluble dietary fibre, *IDF*: Insoluble dietary fibre, *TDF*: Total dietary fiber, *TOS*: Total oligosaccharides

0.7–4.1%, 22.6–63.4 g/kg, 0.0–26.7 g/kg, 4.1–10.3 g/kg, and 11.6–25.4 g/kg, respectively (Tosh and Yada 2010).

Peas are enriched with protein, therefore, used as an essential component in many food product-based businesses. In general protein content in peas varied from 20% to 22%, however, it is reported to vary from 13% to as high as 35% (Irzykowska and Wolko 2004; Tzitzikas et al. 2006). Pea protein has been recognized as a nonallergic foodstuff with excellent nutritional status (Ge et al. 2020). Pea protein is distributed into four distinct classes: globulin, albumin, prolamin, and glutelin, of them, globulin and albumin are the key storage protein that accounts for 55–65% and 18–25%, respectively. Globulin is further differentiated into legumin and vicilin. The other proteins (prolamin and glutelin) are available in very small quantities (3–5%) (Lu et al. 2020). The crude protein of pea comprises of 10–15% of nonprotein nitrogenous stuff and the remaining crude protein (70–80%) is composed of enzymes, hormones, enzyme inhibitors, and nonstorage and storage proteins (He et al. 2020). The pea protein is rich in lysine, leucine, and phenylalanine and poor in sulphur-containing amino acids, i.e., methionine and cysteine (Lu et al. 2020).

Peas are vitamin and mineral-rich legume crops, and their consumption has numerous health paybacks. Amongst the minerals, potassium is the most noticeable mineral with about 1.0% of dry weight, followed by phosphorus (0.39%), magnesium (0.10%), and calcium (0.08%). The occurrence of other minerals such as Fe (97 ppm), Se (42 ppm), Zn (41 ppm), Mo (12 ppm), Mn (11 ppm), Cu (9 ppm), and B (4 ppm) have been reported in pea. The yellow pea seeds possess higher content of Fe, Mg, and Mn, as compared to green peas but are low in K. A substantial amount of variation was reported for Fe, Zn, and Mg which ranged between 46–54, 39–63, and 1350–1427 mg kg^{-1}, respectively. Furthermore, peas have also been considered as an excellent supplement of folate which is higher in green seeded than a yellow seeded pea (Han and Tyler 2003; Jha et al. 2015a). In pea, fat content is low and seeds are an appreciable source of vitamins namely vitamin C, thiamine, pyridoxine, niacin, and riboflavin.

Corresponding to other pulse crops, peas are an excellent source of various phytochemicals namely carotenoids, chlorophyll, β-carotene, and phenolic compounds. The seed coat and cotyledon of colored seeds are enriched with phenolic compounds (Duenas et al. 2004). On a similar note, dark seed coat peas are rich in tannin content, which is having high antioxidant properties. In addition, a subgroup of flavonoids namely Isoflavones also exists in peas in a noticeable amount. The testa and cotyledon of pea seeds are the primary reservoirs of dietary fibre, of them, cellulose (2.4%) predominant in testa while the cotyledon harbored hemicelluloses (1.0%), pectin (1.2–1.7%), and lignin (2.5%) (Tosh and Yada 2010).

3 Nutraceutical Improvement: Cultural and Genetic Biofortification

In addition to the nutritional advantages, the pea is also having certain antinutritional elements like trypsin inhibitors, lectins, oligosaccharides, gallic acid, and other phenolic substances with phytoestrogenic properties. The given compounds could

be either toxic for example lectins, glycosides, alkaloids, indigestible like tannins, saponins, oligosaccharides, or antinutritive like phytates. Such an antinutritional substance diminishes the bioavailability of nutrients in the course of ingestion in the human intestine. The nutritional profiling of staple crops can be uplifted by shrinking the magnitude of antinutritional elements and increasing the concentration of nutritional components by adopting various strategies. The application of orthodox and contemporary breeding procedures to design nutritionally compact varieties is an effective and viable approach to upsurge the bioavailability of minerals. Identification of particular nutritional character(s) is a prerequisite to set objectives to elevate or downsize their level by biofortification. The insufficiencies of Fe, Zn, and Se have been globally noticed particularly among the citizens of underdeveloped nations and approximately 60, 33, and 15% population are experiencing deficiency, respectively (Hotz and Brown 2004). During the last two decades, concentrated efforts were dedicated to conquering the malnutrition problems by employing various tactics such as dietary value addition, food fortification, and biofortification. In biofortification, the nutritive amelioration of the food crops is performed by implication of agronomical techniques and plant breeding/genetics strategies to ensure the supply of adequate essential micronutrients and vitamins in the daily diet of resource-poor populations and vegetarian folks.

3.1 Postharvest Techniques

In the form of dietary supplements beta-carotene has been microencapsulated in pea protein isolate with and without maltodextrin and can be utilized as a worthy component in various food materials, health supplements, and drugs microencapsulation. It can be considered a transitory arrangement to improve the nutritional status of resource-rich people, but it is not sustainable at a large scale. Many food processing/postharvest procedures, viz., dehulling, soaking, germination, autoclaving, fermentation, micronization, and roasting can uplift the lusciousness of pea byproduct along with increment in bioavailability of nutrients by eliminating antinutritional compounds (Ma et al. 2015). Significant changes in physicochemical properties in seeds occur like gelatinization, puffiness of starch, breakdown of protein structure, solubilization of polysaccharides, and softening of structure which transforms the functional properties as well as the deliciousness (Ma et al. 2011). Most interestingly, boiling of frozen peas surges β-carotene level, as cooling and boiling practices break down cell configuration and release the element which was confined with other components in raw peas. Urbano et al. (2006) reported that the sprouting of peas enhanced the bioavailability of Zn and Mg in peas. On the contrary, seed drenching and sprouting for 18 and 48 h decline the level of polyphenols up to 52% and 88%, respectively. Remarkably, seed drenching coupled with the removal of the hull from seed and pressure cooking trickled down the polyphenols by 76% (Bishnoi et al. 1994). Milling and boiling reduce Se content in grains, while improving Zn bioavailability by reducing phytate (Thavarajah et al. 2008). The intake of 100 g of cooked peas biofortified by Se (0.06) and Zn (0.5) is sufficient to meet 50 and 45% of the daily suggested requirement of zinc and

selenium, respectively (Fairweather-Tait et al. 2011). Most recently, different processing methods were employed in pea and a significant increase in the level of protein and starch digestion ability coupled with a decrease in the activity of trypsin inhibitor and level of tannin was observed (Ma et al. 2017). Given reports could be considered as elementary information to explore the opportunities of peas as ingredients, particularly in the agri-food business to boost the competency of peas in terms of augmented nutritional and techno-functional potentials.

3.2 Agronomic Biofortification

In pulse crops very limited efforts have been made toward agronomic bio-fortification. It could be an efficient tactic to intensify the level of Zn and Se in the eatable portions of crops cultivated in Zn and Se deficit fields (Parihar et al. 2021). In agronomic biofortification, seed priming, seed coating, and soil / foliar fertilization are the imperative approaches that can enhance the micronutrient level in the grain without compromising the yield potential. Foliar application of iron and zinc has increased the Fe content in grains under Fe and Zn dearth soils. Interestingly, it has been noticed that blossoming is the appropriate stage for the foliar spray of iron to boost its concentration in grains (Khamparia et al. 2010). Pea has been considered most pertinent to supply Se in the diet due to its selenium accumulation ability in seeds. Foliar spray of 10 g Se per hectare is adequate to upsurge Se levels near to daily recommendation. On a similar note, the foliar application of zinc alone or with soil zinc application enhanced the grain zinc content. The use of zinc sulfate ($ZnSO_4$), zinc oxide (ZnO), ferrous sulphate ($FeSO_4$), and other micronutrient encompassing fertilizers in soil possess excellent promise to escalate crop grain yield and nutrient profile in mature seeds (Shivay et al. 2016). Spraying of Se (0.06) and Zn (0.5) improves the grain nutrient profile in such a way that intake of cooked grains (100 g) could furnish 5%, 35%, and 90% of the daily recommendation of Ca, Mg, and Fe, respectively. Therefore, biofortified peas with iron, zinc, and selenium through the aerial spray of respective nutrients can be the most efficient agro-fortification tactics to combat prevailing malnourishment. To improve the micronutrient level in grains rigorous efforts have to be dedicated to ascertain the potential of a variety of fertifortification tactics. There are several vital components that influence nutrient enrichment of grains including crop, genotype, environment, soil condition, etc., which may also need to be taken into consideration before any agro-fortification recommendation.

3.3 Genetic Biofortification Though Traditional Breeding

Genetic biofortification is offering a more sustainable and economic approach as compared to agronomic biofortification to cater to the requirement of essential micronutrients of developed and developing nations vegetarian folks. Conventional breeding is well acknowledged as the most sustainable and beneficial substitute for

transgenic and agronomic-based strategies to stride against the malnutrition problem across the countries. Unlike cereals, very limited efforts towards biofortification have been made in pulse crops. In the HarvestPlus program ICARDA has released numerous high-yielding varieties of lentil with high iron and zinc suitable for different ecologies. On a similar note, using conventional breeding four cowpea improved cultivars namely PL-1, PL-2, PL-3, and PL-4 have been released with high iron content under the HarvestPlus program. In the case of common beans, to date, 10 varieties with improved Fe and Zn content are developed in Rwanda and Congo in association with the HarvetPlus platform. Unfortunately, in peas very limited efforts are dedicated toward bio-fortification. Nevertheless, it is inherently well-off in Fe, Zn, Mg which oscillated from 46–54, 39–63, and 1350–1427 mg kg^{-1}, respectively. In another report the Fe, Zn, Ca, Mg, and Se levels oscillated between 45.2–48.9, 32.3–35.0, 786–802, 1210–1270, and 0.413–429 mg/kg, respectively, in Canadian conditions. Protein is one of the vital compounds present predominantly in pea seeds in addition to starch and fibers. Some of the investigation findings have illustrated ample variability of protein content in pea that fluctuates between 14% and 38% (Tzitzikas et al. 2006). To quantify protein various tools and techniques, i.e., Kjeldahl, Calorimetric, NIR, Dumas, and amino acid analysis have been adopted that showed substantial variation for protein content among the studies. In a report comprising of 1146 pea accessions, the protein content ranged from 13.2% to 30.9% with a mean of 24.4% in the Genetic Resource Information Network (GRIN) program (USDA 2020). Some improved cultivars, viz., Cameor, VavD265, and CDC Striker were evaluated for protein content in different studies, and these cultivars demonstrated high protein content (>26%). However, in addition to genetic factors, environmental conditions also play a major role in the variation of protein content. Besides, enrichment of grain micronutrient status, the bioavailability of nutrients needs to be increased through conventional breeding by reducing anti-nutritional elements and escalating compounds that encourage iron assimilation. For example, phytate declines the bioavailability of micro-nutrients; therefore, the reduction in phytate level is the utmost tactic toward the biofortification of pea. The biochemical pathway of phytate was modified by mutagenesis followed by conventional breeding. In pea, two low phytate mutants, viz., 1-150-81 and 1-2347-144 were developed employing chemical mutagenesis in base parent CDC Bronco (Warkentin et al. 2012). To understand the effects of environmental conditions on phytate level, multilocation testing was performed and results revealed that inorganic phosphorus, phytate phosphorus, and level of iron were significantly changed over the locations (Warkentin et al. 2012). In another study, it has been found that the reduced phytate in pea is operated through a recessive single gene (Rehman et al. 2012). The carotenoid enriched peas seeds are part of a biofortification approach and recently reported that the green-colored cotyledon pea cultivar possesses twice the total carotenoids as yellow cotyledon cultivar. Notably, the genotypic effect influences carotenoid content to a greater extent than the environment. The association analysis revealed that iron content has a positive correlation with iron bioavailability, while phytate content has a negative correlation with iron bioavailability. In addition, lutein content has a positive correlation with iron bioavailability.

The conventional breeding approach has been well accepted as it is cost effective, free of synthetic inputs, eco-friendly, and sustainable practice. In spite of the several benefits of employing traditional breeding to bio fortify foods, certain limitations are also there; the first one is dependency on the existing genetic diversity for the trait of interest in targeted crop gene pools, i.e., primary, secondary, tertiary. If substantial diversity is not available then biofortification through conventional breeding in the targeted crop is not possible. For instance, biofortification in oilseed could be possible only through transgenic approach owing to its narrow genetic base, low heritability, and linkage drag with targeted traits. Furthermore, nutritional profiling of contemporary crops and crop wild relatives (CWR) has demonstrated that some of the crops are poor in nutritional worth than their wild colleagues. Though, the exploitation of CWR in biofortification could be a difficult task in owing to their cross incompatibility, tight linkage with undesirable traits with targeted traits and least representation of CWR in world gene bank. Another important drawback of conventional breeding is that it takes years to develop a cultivar because incorporation of a trait into an elite cultivar has to go through an exhaustive selection procedure at least till sixth generation. Conversely, many approaches are there like high-throughput phenomics platforms, seed chipping technology, molecular markers, genomic selection, genome editing, and speed breeding that can hasten the varietal improvement process. Of which, some of the methods are expensive than traditional breeding methods, therefore, not used extensively in public breeding institutes albeit they are more efficient. Last but not least, in conventional breeding before releasing, the cultivar must be tested in different environments because genotype-by-environment relationship can considerably affect cultivars phenotypic and nutritional performance. Consequently, a biofortified variety may lose its improved nutritive trait owing to the genetic-by-environment interactions.

Laborious and time-consuming nature of conventional breeding and availability of genomic tools prompted plant breeders to go for marker assisted selection. Biotechnological approaches have played an instrumental role in crop improvement and could be helpful in designing food crops with targeted nutritional profile. The amalgamation of conventional breeding with biotechnological intervention has resultant improved cultivars with high nutritional profile. The most important activity for the detection and development of molecular markers associated with micronutrient is précised phenotyping of available germplasm accessions and successive discovery of contrast genotypes for trait of interest, which will be used to develop appropriate experimental populations. Mapping populations will be exploited to identify molecular markers related with the trait of interest. Ultimately, the identified genetic variants accountable for an amplified level of the targeted nutrient should be introgressed in targeted genotype to emanate the nutrient rich cultivar, either by traditional breeding or by current biotechnological tools. Genomic resources let breeders to take advantage of the existing genetic variability more precisely; hence, duration and expenses could be reduced significantly. In traditional plant breeding, the selection is carried out just considering phenotype, however, by applying genetic and genomic apparatus, allelic variants can be allied with phenotypic variation, which allow early selection of the plant. In the recent past, efforts

were dedicated in pea towards development of genomic resources (Krajewski et al., 2012). Unlike other legume crops, in pea genomic resources have not been exploited judiciously in order to identify genes or quantitative trait loci (QTLs) for nutritional traits. However, some of the researchers revealed the genetic mechanism of iron content in seeds and successfully obtained markers and quantitative trait loci to facilitate breeding programmes (Gali et al. 2018). Most recently, SNPs markers associated with iron and zinc were identified using a group of genotypes of pea (Diapari et al. 2015). On a similar note, SNPs marker allied with Ca, Mg, carbohydrates, and inositol were found in pea (Cheng et al. 2015). Besides, QTLs were also identified on LG3, LG4, and LG7 for Fe status in a RIL population derivative of Carrera/CDC Striker. By employing genome wide association approach in a recombinant inbred line (RIL) population of Kiflica/ Aragorn, many QTLs have been discovered for seed weight and mineral concentration (Ma et al. 2017). Biotechnological approaches have been utilized to support breeding programs, such as marker-assisted selection (MAS) which dramatically accelerated the achievement of breeding for biofortification program. Unfortunately, pea being oldest domesticated crop is still lagging behind in the case of genomic resources. To accelerate genomics enabled biofortification there is urgent need of creation of more genomic resources and their judicious utilization towards identification of genes/QTLs that will be an important asset in MAS. The brief description of markers and molecular mapping of nutrient related genes and QTLs is presented in Sect. 6 of this chapter.

If genetic variability is inadequate for trait of interest then the desired variability may be created by adopting induced mutagenesis or modern techniques like genome editing (i.e., clustered regularly interspaced short palindromic repeats (CRISPR) – associated system (CRISPR/Cas). The beauty of genetic engineering is that it can utilize indefinite group of genes for the targeted transfer and expression of desirable characters from one organism to another without any taxonomic constraints. Besides, if a targeted nutrient is not synthesized in a particular crop, to biofortify such crops for that specific nutrient transgenic is the best option. Using transgenic approach genes were integrated in the genome of the targeted crop to manufacture the micronutrient, for instance, golden rice. Initially, transgenic technique involves considerable duration, labor, and expense, but in a long term, it is lucrative and sustainable unlike other biofortification approaches. On this line biofortification efforts are underway in several crops including pea and brief account of transgenic and genome editing is elaborated in respective section of this chapter.

4 Pea Genetic Resources

4.1 Current Germplasm Holding

Pea is considered as the world's oldest domesticated crop and is cultivated globally in temperate climates. It has been cultivated from Neolithic period as a food and vegetable crop. Pea is the fourth largest pulse crop cultivated mainly for a pulse,

fodder, and vegetable. As per the information in the GENESYS portal, total collections of pea amounted to 33,023 accessions, conserved in different genebanks in the various parts of the world (Genesys portal, accessed on August 30, 2022). United Kingdom has the highest holding of germplasm collections followed by Australia, India, and Lebanon. The list of major genebanks holding pea germplasms are given in Table 2.

Table 2 Holdings of pea collections in different genebanks

S. No	Holding Institute		Country of holding institute	Total collections
	Gene bank code	Total collections		
1	GBR247 GBR165 GBR016 GBR017 GBR006	3360 3298 2110 132 02	United Kingdom	8902[b]
2	AUS165	7325	Australia	7325[b]
3	VIR	6790	Germany	6790[c]
4	USDA	5400	United States of America	5400[d]
5	NBPGR[a]	4680	India	4680[a]
6	LBN002	4596	Lebanon	4596[b]
7	ITA436 ITA394	1716 1225	Italy	2941[b]
8	UKR001	2704	Ukraine	2704[b]
9	SWE054	1749	Sweden	1749[b]
10	BGR001	1570	Bulgaria	1570[b]

GBR247-Germplasm Resources Unit, John Innes Centre. Norwich Research Park – NR4 7UH, Norwich, United Kingdom
GBR165-EURISCO (European Search Catalogue for Plant Genetic Resources) – https://www.sasa.gov.uk/
GBR016-EURISCO (European Search Catalogue for Plant Genetic Resources) – https://www.igergu.ibers.aber.ac.uk/
GBR017-EURISCO European Search Catalogue for Plant Genetic Resources https://www.gardenorganic.org.uk/
GBR006-EURISCO European Search Catalogue for Plant Genetic Resources http://www2.warwick.ac.uk/fac/sci/lifesci/acrc/gru
AUS165-Australian GrainsGenebank, Agriculture Victoria, Australia
LBN002-International Centre for Agricultural Research in Dry Areas
ITA436-Consiglio NazionaledelleRicerche – Dipartimento di Scienze Bio-Agroalimentari, Italy
UKR001-Institute of Plant Production n.a. V.Y. Yurjev of UAAS, Ukraine
SWE054-Nordic Genetic Resources Centre (NordGen). Växthusvägen 24–23,456, Alnarp, Sweden
BGR001-Institute of Plant Genetic Resources "KonstantinMalkov". StrDrujba 2–4122, Sadovo, Plovdiv district, Bulgaria
[a]NBPGR- National Bureau of Plant Genetic Resources, New Delhi, India-http://www.nbpgr.ernet.in/Research_Projects/Base_Collection_in_NGB.aspx-. Accessed on August 30, 2022
[b]Source: https://www.genesys-pgr.org/-. Accessed on August 30, 2022
[c]VIR-N.I. Vavilov Research Institute of Plant Industry, St. Petersburg, Germany
[d]USDA-Plant Germplasm Introduction and Testing Research Station, Pullman, United States of America

4.2 Primary Genepool

The genepool concept was given by Harlan and De wet to categorize the hybridization relationship of a species with other species and grouped into three categories, primary, secondary, and tertiary genepool. However, advancement in plant genomics has led to another set of category defined by Hammer and Gepts and Papa as a fourth genepool (quaternary genepool), and it could include any synthetic strain revealed in nucleic acid sequence, DNA or RNA that do not exist in nature. According to this concept, the species which are in the primary genepool of a crop can be easily crossable and develop fertile progenies during hybridization. It includes plants of the same species or closely related species wherein genes tradeoff can be done between the species by simple crosses, and it is considered as the précised material in term breeding importance. It is believed that *P. sativum* was domesticated in the Near East about 11,000 years ago, likely from *P humile* (also known as *Pisum sativum* subsp. *elatius*). In the *Pisum* genus, the species which are in the primary genepool are *Pisum sativum* subsp. *elatius*, *Pisum sativum* subsp. *elatius* var. *brevipedunculatum*, *Pisum sativum* subsp. *elatius* var. *elatius*, *Pisum sativum* subsp. *elatius* var. *pumilio.*

4.3 Secondary Genepool

The species in the secondary genepool results partial fertile hybrids on crossing with primary genepool. It includes plants related to the species. The introgression of gene from such material to primary genepool is feasible but difficult. The *Pisum* species which comes under this category are *Pisum abyssinicum* and *Pisum fulvum.*

4.4 Tertiary Genepool

The species in this category will lead to the synthesis of sterile hybrids on crossing with primary genepool, and the transfer of genes between the species is possible only with the help of exceptional techniques like bridge crossing, genetic recombination, etc. The species of peas which are under tertiary genepool are *Vavilovia formosa.*

4.5 Sources of Donor Genes

The proper utilization of genetic resources in any breeding program mainly relies on a valorization of genetic resources for targeted traits. In case of evaluation of existing germplasm, accessions for quality traits pea is still lagging behind other legume crops. During last 3–4 decades, concerted efforts have been made to characterize the pea germplasm for various nutritional traits, i.e., protein, iron, zinc, RFOs, and antinutritional traits (Tzitzikas et al. 2006; Tosh and Yada 2010; Harmankaya et al. 2010; Dahl et al. 2012; Demirbas 2018). Most importantly, the landraces and CWR could serve as arsenal of genetic resources to breed new crop varieties suitable for

changing environmental and demographic conditions. Legume crop holds an exceptional status in world agriculture being protein rich and inbuilt ability of environmental nitrogen fixation, but bioavailability of protein is a major concern, because, the protease inhibitors decreases protein digestibility and are operated by two genes (TI1 and TI2) in pea. Recently, a wild pea (*P. elatius*) accession got mutated within the aforesaid both genes and considerably decreased the scale of protease inhibitor activity improving the bioavailability of amino acid (Clemente et al. 2015). Gawłowska et al. (2017) evaluated about 248 accessions belonging to *P. abyssinicum*, *P. elatius*, *P. fulvum*, *P. syriacum*, *P. sativum* subsp. *asiaticum, P. sativum* subsp. *transcaucasicum, P. sativum* subsp. *sativum* convar. *Axiphium*, convar. *Medullosaccharatum* and *medullare*, convar. *Vulgare* and convar. *Speciosum* for RFOs. The results revealed that highest content of total soluble carbohydrates and total RFOs were noticed in accessions with wrinkled seeds and the lowest content in the wild species *P. fulvum*. Therefore, *P. fulvum* could be used as valuable source in conventional breeding to further decrease RFOs. Most recently, in another study Ethiopian pea (*Pisum sativum* var. *abyssinicum)* landraces were examined to estimate the nutritional composition, and results demonstrated abundant variation for nutrients and mineral content. The protein content varied from 21.63% to 28.13%. Furthermore, all the landraces had high potassium (41.43–74.21 μg/g) and low sodium (0.93–27.65 μg/g) content (Gebreegziabher and Tsegay 2020). These finding clearly indicated that Ethiopian pea is an excellent source of protein, and other minerals and could be utilized in regular breeding to develop cultivars with enhanced nutritional status. In addition, *P. fulvum, P. sativum* subsp. *elatius* var. *pumilio*, *P. abyssinicum,* and *P. sativum* subsp. *elatius* also provide resistance against many biotic and abiotic stresses like rust, pea weevil, powdery mildew, ascochyta blight, broomrape and bruchid, bacterial blight, frost tolerance. *P. sativum* subsp. *sativum* var. *arvense* has potential to be utilized as feed and fodder for livestock. The given genotypes could be used in conventional and molecular breeding to develop high yielding and nutritionally enriched cultivars, experimental populations and established marker trait association to identify genes/QTLs. Henceforth, intensive efforts should be made toward the evaluation of available germplasm accessions for various nutritionally important traits, which could be used in breeding program to develop nutrient dense cultivars that would be helpful in the improvement of the overall health status of resource poor and vegetarian population of developed and developing countries.

5 Genetic Mechanism of Nutritionally Important Traits

Pea protein is complicated genetically because of various multigenes families encrypting protein composition and content. Pea seed protein composition is operated genetically; however, environmental factors and postharvest processing circumstances also influences protein magnitude (Bourgeois et al. 2009; Bourgeois et al. 2011). An earlier report identified that three QTLs accounting for 45% of the total variation of protein (Tar'an et al. 2004). On a similar note, 14 QTLs were identified

across the environment affecting the development of genes controlling pea seed protein concentration (Burstin et al. 2007). Another report in which a RIL population was used and detected significant QTLs on linkage group V for total protein content (Irzykowska and Wolko 2004). Similarly, in two populations synthesized by hybridization between a small-seeded and protein rich line and two diverse large-seeded and high-yield potential cultivars detected QTLs for seed protein content located on LG V (Krajewski et al. 2012). Recently, a QTL on LG VI was detected that demonstrated 45% of the seed protein content variation. In addition, single-nucleotide polymorphism (SNP) within *O_2like* gene in pea was identified having a significant consequence on seed protein content. *O_2like* genes in pea have resemblance to *Opaque2* (*O2*) of maize, a bZIP transcription regulating factor that influences starch and protein content (Jha et al. 2015b).

The pea seed protein largely comprises of globulins legumin and vicilin storage proteins, where its synthesis is operated by at least 40 genes and 10 different genetic loci (Casey et al. 2001). Nevertheless, as mentioned above pea seed protein content as quantitative trait, therefore, single gene mutations may be less efficient to influence total protein concentration if not the mutation is at a locus which encrypt substantial part of phenotypic variation (Bourgeois et al. 2009; Rayner et al. 2018). For legumin biosynthesis up to four genes which are defined by distinct loci have been reported and approximately there are 10–15 genes which produce protein (Casey et al. 2001). Interestingly, the loci control starch synthesis can also influence the legumin synthesis due to mutant alleles at loci (*r* and *rb* controlling starch synthesis and synthesis of large subunit of ADP-glucose pyrophosphorylase, respectively) influence the expression of legumin genes and ultimately manipulates the ratio of legumin/vicilin in the overall seed protein (Casey et al. 2001).

Another important storage protein is vicilin that is a poor source of sulphur-containing amino acids in comparison to legumins. Consequently, these amino acids designated as negative factor in pea, and increment in their level is considered as a prime objective toward the nutritional upgrading. In cotyledon, vicilin produces amyloids which are protein aggregates having unique physiochemical attribute like resistance to alimentary tract digestion (Antonets et al. 2020). Using proteomics technique, 24 genes were found controlling vicilin synthesis in pea and also develop a reference map of pea seed proteome (Bourgeois et al. 2009). Interestingly, Domoney et al. (2013) demonstrated how the fast-neutron mutagenesis at single loci knocks out multiple genes which could influence synthesis of vicilin and some other proteins. On a similar note, a mutant allele at *Vc*-2 position influenced the manufacture of vicilin polypeptides and that subsequently impacted seed protein concentration (Chinoy et al. 2011). In a QTL analysis, genes operating legumin, convicilin and vicilin have been identified (Bourgeois et al. 2011). In another investigation, wherein characterize loss of function of *abi*5 mutant in pea and noticed that vicilin content reduces in mutant, and genes encrypting other major protein were upregulated (Le Signor et al. 2017). The gel permeation chromatography demonstrated the proportion of legumin and vicilin in total protein predominantly remains under genetic control, however, agronomic components also influences (Mertens et al. 2012). Most recently, reference genome of pea has been available which offer great opportunity for seed protein

annotation. The genome assembly was explored, and DNA sequences programming proteins counting legumin, vicilin, and convicilin have been recognized. In the pea reference genome alleged RY motifs (CATGCATG) up regulated three storage protein genes by regulating their expression (Kreplak et al. 2019).

There are two proteins, i.e., lectin and albumin which are undesirable due to poor digestibility and allergic effects (Le Gall et al. 2007). In earlier reports it is mentioned that, the lectin is determined by *LecA* gene that control symbiotic interactions of roots with rhizobium (Díaz et al. 1989). Most recent study illustrated the nodulation competence of the *LecA* mutant and observed that it is similar to that of wild-type pea, advocating that absence of lectin is not affecting the symbiotic interaction (Rayner et al. 2018). For pea albumin 2 (PA2) and PA1 coding nine and eight gene sequences, respectively, could identify using the reference genome (Kreplak et al. 2019). In another study, the null mutant for PA2 was hybridized with cultivar (Birte), and a RIL population was synthesized. The examination of RIL exhibited that the lines deficient to PA2 were higher in seed nitrogen content probably owing to a surprising upsurge in the production of other seed proteins (Vigeolas et al. 2008). Backcrossing with cv. Birte has also revealed the PA2 function in plant growth and response to stresses (Vigeolas et al. 2008). This was further validated in a report wherein it is explained that the polyamine and spermine can unite to the homolog of PA2 in grasspea (Gaur et al. 2010).

Two major Lipoxygenases (LOX) polypeptides are present in pea seeds, viz., LOX-2 and LOX-3, of which, one catalyze the creation of compounds which are not preferred by customers (Forster et al. 1999). Both polypeptides are programmed by two or three genes at the *lox* locus situated on LG IV (North et al. 1989). However, a null mutant for LOX2 has been identified from *P. fulvum* collection at the John Innes Centre. Further analysis of this locus in this particular accession portrayed that the decrease in LOX-2 mRNA was not the outcome of a loss of LOX genes, but it happened owing to the one or multiple insertions, deletions, or substitutions into the LOX-2 promoter of the null mutant (Forster et al. 1999). Being rich in sulphur-containing amino acids, pea albumins are vital component for human, and questions are often raised over low level of these amino acids in legumes. Hence, pea albumins can offer a nutritional advantage in respect to amino acid configuration, notwithstanding, apprehension are also there regarding its allergic properties and resistance to ingestion of these proteins in human guts.

Starch, being a most dominating storage carbohydrate in pea seeds accounts for about 50% of dry seed weight. It is composed of various segments considering their configuration and digestibility (Lockyer and Nugent 2017). In his experiment Mendel used seven external characteristics of pea, including variation in mature seed shape, i.e., round or wrinkled to established laws of inheritance (Bateson 1901). Initially, this trait or locus was named as *rugosus* (gene symbol "r") by white (1917) derived from Latin for wrinkled or shriveled shape and is located on chromosome 7 (Bhattacharyya et al. 1990). Later on, another locus named *rb* produces wrinkled shape seeds was recognized (Kooistra 1962). The *r* locus conferred a prominent visible effect on seed appearance that resulted due to the dominant effect of this locus on the configuration of pea seed. There is an unambiguous variation in starch

metastasis of round and wrinkled seeds of peas (Bhattacharyya et al. 1993). In addition, round seeded pea contain more starch than wrinkled seeded and elevated proportions of amylopectin to amylase. The *rr* seeds possess greater concentration of free sucrose than round seeded, which leads to the higher osmotic pressure and water content and bigger cell size of wrinkled seeds (Wang et al. 1987). The *rr* seeds lose a greater part of their volume in the course of seed development, and because the testa does not contract along with the cotyledons, it shrinks to produce the wrinkled appearance (Casey and Davies 1993). Later on the wrinkled phenotypes were also reported to occur due to mutation at other loci counting *rb* and *rug* 1, 2, 3 (Wang and Hedley 1991). Notably, all the mutant, viz., *r*, *rb,* and *rug* have analogous exterior appearance, however, *r* mutants can be discriminate based on the starch granule morphology (Bhattacharyya et al. 1990). Starch granules are build up of two types of starch polymers, i.e., amylase and amylopectin. The proportion of amylose in the wrinkled seeds is around twofold of round seeds (Bhattacharyya et al. 1990). Indeed, the wrinkled seeds developed due to mutation in the gene that governs starch-branching enzyme isoform I (*SBE1*) by inclusion of a transposon-like component into the coding arrangement. Two isoforms of starch-branching enzyme, viz., *SBE1* and *SBEII* are expressed in the initial and later phase, respectively, of embryo development. *SBEI* accounted about 1/3 of the amylopectin in matured pea embryos and create a less soluble polymer in comparison to the polymer generated by *SBEII* which may be more effective in catalysis of short chain production (Burton et al. 1995). The mutations of *SBEI* significantly reduce the scale of starch and fraction of amylopectin in the wrinkled genotypes than round (Bhattacharyya et al. 1993). The reduction in starch synthesis and failure of amylase to amylopectin conversion mainly happened due to the loss of SBEI enzyme activities. The metabolic proof and decreased SBE activity clearly designated this enzyme as an imperative determinant in starch content of *rr* embryos (Bhattacharyya et al. 1993). The SBEI play essential role in the construction of usual starch granules that cannot be replaced by other isoforms. In addition, *r* mutant seeds also demonstrated pleiotropic effects and intricate metabolic instabilities, for example, elevated scale of free sucrose, extra lipid, less legumin, and subsidize seed longevity (Bhattacharyya et al. 1990). In *SBEII,* gene mutation has not been recognized by the examination of wrinkled seeds which advocated that the mutation of SBEII does not have any role in wrinkled phenotype, and this could be due to minor involvement of SBEII in amylopectin synthesis and subsequent activity in the course of embryo development (Bhattacharyya et al. 1993). Starch debranching enzymes (DBEs) hydrolyze the α-1,6-glucan branches of amylopectin, this step is essential for the regular synthesis of amylopectin by pruning extra branches (Wang et al. 2014). In fact, two categories of debranching enzyme are there, viz., isoamylase (ISA) and pullulanase (PUL) with different amino acid sequences. Both DBEs are present in peas and have important role in starch metastasis during the development of embryo and in starch breakdown in the course of germination (Zhu et al. 1998). During last decade, three genes coding the diverse isoforms of ISA (*Psisa1*, *Psisa2,* and *Psisa3*) have been identified in peas and these are capable to bind to glucan substrate, although *Psisa2* deprived of catalytic skill (Hussain and Martin 2009).

At the end of the twentieth century, four other loci were recognized which influences starch metastasis: *lam, rug3*, *rug4*, and *rug5*. Mutant of *rug4* locus causes loss of sucrose synthase (*Susy*) activity during the development of pea embryo. The *rug4* mutant develops wrinkled seed and reduced the starch level in the embryo by affecting the supply for starch biosynthesis (Craig et al. 1998). Another locus *rug5* control starch synthase II, and mutation results into modification of starch granule morphological appearance and the configuration of amylopectin (Craig et al. 1998). The *lam* mutation also influences starch synthase enzyme (starch synthase I) and develop starch with low amylose and high amylopectin (Tahir et al. 2011).

Genes governing three isoforms of Susy (*SuSy1*, *SuSy2*, and *SuSy3*) were identified in peas. The starch scale of the *SuSy1* embryo is reduced by 30%, whereas the cellulose level remains constant. The *SuSy1* isoforms in embryos is essential for starch biosynthesis but not required for synthesis of cellulose (Weber et al. 1998). In pea, *rug3* locus is responsible for a plastidial phosphoglucomutase (PGM) enzyme that catalyzes the interconversion of glucose-6-poshpate (Glc-1-P) and Glc-6-P in the cytosol and plastids. Further, five pea mutants at the *rug3* locus developed wrinkled seeds with low amylose and starch content. In pea embryos, Glc-6-P is ingress from the cytosol into the amyloplast where it is reconverted to Glc-1-P by the plastidial isoform of PGM by supplying the substrate for starch synthesis (Harrison et al. 1998). The damage of plastidial phosphorylase (Pho1) enzyme causes substantial decline in the starch magnitude and its configuration (Satoh et al. 2008). Other two *Pho* isoforms, i.e., *Pho1* and *Pho2* were identified in pea cotyledon, of which Pho2 plays a vital role in starch granule development (Van Berkel et al. 1991). Recent literature reported auxin to play an imperative role in regulating starch accrual in pea seeds (McAdam et al. 2017). The auxin deprived mutant *tar2* (tryptophan aminotransferase related 2) develop small and wrinkled seeds with low starch level. The activity of different starch synthesis enzymes and expression of the corresponding genes are condensed in mutants indicative of vital role of auxin in starch accumulation in peas (Meitzel et al. 2021).

The above given facts witnesses the magnitude and composition of starch as polygenic traits, and their overall expressions are environment sensitive. Genetic mechanisms of starch metastasis are well explained in cereal crops, while there is little information in case of legumes. Recently, 132 SNPs were detected within the genes involved in carbohydrate metabolism, of which, 4 SNPs were associated with genes *AGPase_L1*,*GBSSI*, and *SBEII* operating amylose concentration, and 10 SNPs were associated with genes *SBEII, SuSy2,* and *Sps* (Jha et al. 2015b). Partial sequences of 25 candidate genes which represent 16 enzymes concerned to starch metastasis were characterized using a group of 92 pea accessions, which revealed 3 candidate genes (*r*, *UGPase*, *AGPS2*) to be associated with amylopectin chain length distribution (CLD), while amylose level has association with the *r* locus (Carpenter et al. 2017). Most recently, a linkage map has been developed for starch scale using two RIL populations and detected QTL for starch magnitude on LG2b and LG4a in one population, while QTL for starch content on different linkage groups, viz., LG1a, LG3b, LG3c, and LG7a in other population (Gali et al. 2018). The GWAS has been performed for seed starch content in a panel of 135 pea

accessions and subsequently identified 9 SNP markers positioned on LG2, LG3, LG5, and LG7, and one scaffold was associated with starch level (Gali et al. 2019). Genes encrypting raffinose (*rfs*) and stachyose synthases (*sts*) were aligned to LG III and V, respectively (Ellis et al. 2018). Gali et al. (2018) explored three RILs and identified two QTLs for acid detergent fibre (ADF) positioned on LG 4 and two QTLs on LGVII. In another population, QTLs for ADF have been detected on LG IV and LG VIIa. For neutral detergent fibre (NDF), QTLs on LG Va, Ia, IV, and VIIa have been detected. Similarly, in another study five SNPs associated with level of ADF have been identified on chromosomes 5, 6, and 7, and eight on chromosomes 2, 3, 5, 6, and 7, of which, two markers linked with both type of fibre were positioned on chromosomes 6 and 7 (Gali et al. 2019). Further analysis of identified QTLs and discovery of new QTL could pave the way towards improved discerning of the genetic machinery operating the starch metabolism in pea seeds. Genes associated with starch synthesis are thoroughly examined in pea, consequently offers worthy opportunities for the manipulation and modification of pea starch magnitude and configuration considering further scope. In addition, it has been noticed that the total starch level reduces in peas having high protein, and high amylose level increases resistance to ingestion of the starch (Shen et al. 2016).

Number of attempt has been made to decipher the genetic foundation of seed iron level and consequently detected various molecular marker and QTLs to facilitate breeding programme. Most recently, five QTLs have been detected which are linked with seed iron content in pea localized on LG VII and II (Ma et al. 2017). Of them, three QTLs appeared in close vicinity to markers earlier associated to iron concentration in a study wherein a set of 94 accessions was used (Diapari et al. 2015). QTLs for seed iron content were also reported in three RILs which were derived by hybridization among European and Canadian cultivars (Gali et al. 2018). QTLs for iron concentration in seed were identified on LG IIIb, however, on linkage group VII none of the QTLs for iron content were noticed (Diapari et al. 2015; Ma et al. 2017). This might be owing to the diverse parents used in mapping population or the derived material was evaluated under multilocation testing (Gali et al. 2018). For rest of the essential micronutrients like zinc and selenium very limited efforts were dedicated to understand genetic foundation of their concentrations in pea seeds. However, two SNPs were identified to be linked with high zinc concentration on LGIII (Diapari et al. 2015). Similarly, one more study also reported a QTL on LG III which elucidating maximum magnitude of phenotypic variance for zinc content in seeds (Ma et al. 2017); nevertheless, the SNPs identified by Diapari et al. (2015) were not in adjacent to the QTL located on LG III. A recently conducted study identified four QTLs concerning to zinc level in seeds along with a noteworthy locus on LG III (Gali et al. 2018). Overall, the above narrated finding suggested that the pea genomic region on LG III have huge potential towards the escalation of zinc content in seeds and identification of the core genes/genomic regions that would be helpful for accelerating biofortification programmes in pea. On a similar note, selenium content in seeds was examined in three RILs and of which, in two populations several QTLs located on LG VII, IV, and Vb were detected and strong environmental effect was also observed (Gali et al. 2018). The chemical

resemblances between selenium and sulphur demonstrated that the QTLs recognized in this report may be loci that encrypt sulphate transporters (White 2016).

6 Molecular Mapping of Health-Related (HR) Genes and QTLs

Understanding the genes and genomic location governing valuable trait requires the comprehensive study of whole-genome sequence data. Pea genome consists of seven pairs of chromosomes with a huge genome size of 4.45 Gb (Gali et al. 2018), mostly comprised of transposons and other mobile elements of Ty3/gypsy family (Macas et al. 2007). The generation and accessibility of genomic tools have been slowed down in pea owing to the late availability of reference genome in public database. With the advancement of cutting-edge tools and cost effectiveness of Next Generation Sequencing Techniques (NGS), momentum jump has been achieved towards marker development and subsequent molecular breeding efforts for enriching pea genomic resources. Pea researchers walked a long way with the partnership of International-National consortium for making available the complete pea genome sequencing platform and paved the way from conventional to molecular breeding era (Madoui et al. 2015).

A molecular marker is a segment of DNA linked to a particular locus in the genome. Molecular markers facilitate to perform MAS for the trait of interest at early growth stage and accelerate genetic progress, thus expedite genetic mapping and QTL analysis. Development of first and second-generation markers like RFLP, RAPD, AFLP, SSR, and CAPS aided for diversity studies and pea genetic improvement towards gene tagging and introgression (Aubert et al. 2006; Burstin et al. 2007; Bourgeois et al. 2011; Ma et al. 2017). With the availability of high throughput SNPs, genotyping has become more efficient, thanks to technological improvements in sequencing and generating genotyping platforms over the past 10 years in pea. Presently SSRs and SNPs markers are the prime choice of the plant breeders owing to their reliability, robustness, and wide distribution throughout the genome. Several SNP arrays have been generated by deploying genotyping-by sequencing (GBS) approach and transcriptome data of diverse pea germplasm set (Kaur et al. 2012; Tayeh et al. 2015a; Alves-Carvalho et al. 2015). Genetic map construction in pea dated back early in 1925 through development of six linkage group, however markers development in relation to HR genes were scarce (Table 3). Mapping efforts concerning seed protein quality was initiated to detect the QTL (Tar'an et al. 2004; Bourgeois et al. 2011; Tayeh et al. 2015b). Three QTLs were detected that was associated with seed nitrogen/protein content (Tar'an et al. 2004). Among these, two QTLs were mapped on LG-III and VI and can explain 45% of the total variation in seed protein and likely to be associated with seed storage protein albumin. Another study detected a total of 14 QTLs explaining 9–46% variation, positioned at 160 cM and 170 cM on LG V, thus highlighted the presence of two QTLs with opposite effects (Burstin et al. 2007). Except for the fabatin-like genes on linkage group VII, none of the QTLs were directly linked to seed protein gene loci. Irzykowska and

Table 3 Consensus mapping strategies concerning HR genes in Pea

Trait	QTL /LGs	Mapping Population	Marker system	Map distance (cM)	References
Protein	03 QTLs (LG-VI))	88 RILs between 'Carneval' × 'MP1401'	193 AFLPs 13 RAPDs 01 STS (sequence tagged site) marker	1274	Taran et al. (2003)
	5 QTLs (LG-II, V, VII)	114 plants of F_2 generation from a cross between Wt10245 × Wt11238	204 markers morphological, isozyme, AFLP, ISSR (Inter Simple Sequence Repeat), STS, CAPS and RAPD markers	2416	Irzykowska and Wolko (2004)
	14 QTLs (LG-I, III, IV, V, VI, VII))	139 RILs between Térèse' and K586	249 SSRs	1113	Burstin et al. (2007)
	77 QTLs (LG-V) and 11 LGs	F_2 generations derived from Wt11238 × Wt3557 Wt10245 × Wt11238	282 morphological, isozyme, AFLP, ISSR, STS, CAPS, and RAPD markers	853	Krajewski et al. (2012)
	7 LGs	RIL population from the following crosses: VavD265 × Ballet (211 individuals) Cameor × Melrose (120 individuals) Kazar × Cameor (84 individuals) Kazar × Melrose (118 individuals)	SNPs, SSRs, RAPDs, RFLPs AFLPs, etc. 6188 8503 7013 3917	794.9	Tayeh et al. (2015a)
Vicilin	7 QTLs (VicB, Vc-2-5 (LG II, III, V)	157 RIL population between 'Cameor', 'VavD265', and 'Ballet'	Seed protein markers	–	Bourgeois et al. (2011)

(continued)

Table 3 (continued)

Trait	QTL /LGs	Mapping Population	Marker system	Map distance (cM)	References
Starch metabolism	7 LGs	Térèse × K586 (Pop1, 139 F7 RILs) Térèse × Champagne (Pop2, 164 F8 RILs)	63 SNPs and 15 Indels markers, CAPs, and RAPDs	1458	Aubert et al. (2006)
Seed mineral content	82 QTLs	RIL population (F6) derived from 'Kiflica' and 'Aragorn'	114 SSRs and 1608 SNPs	1310.1	Ma et al. (2017)
Dietary fibre, seed protein, minerals, and phytate acid	16 QTLs (LG-Ib and LG4a: Seed protein) (LG-IV, VII and VIIa: acid detergent fibre) (LG-Ia, IV, Va, and VIIa: neutral detergent fibre (LG-IIb and LG-IVa, LG-Ia, LG-IIIb, LG-IIIc, and LG-VIIb: seed starch content) (LG-IIIb: seed Fe content) (LG-VII, LG-IV, and LG-Vb: seed Se content 4 QTLs for seed Zn content (two on LG-VI; one each on LG-Ia and LG-IIIb) (LG-IIIa, LG-V, and LG-VIa: seed phytate content)	RIL population developed from PR-02 (Orb × CDC Striker), PR-07 (Carerra x CDC Striker), and PR-15 (1–2347-144 × CDC Meadow)	2066, 3023, and 3444 SNPs for three different population	951.9, 1008.8 and 914.2 cM for three different populations	Gali et al. (2018)
Phytic acid and phosphorus content	7 LGs	RIL population with 94 individuals derived from cross between 1–2347-144 × CDC Meadowe	341 SNPs	771.6	Sindhu et al. (2014)

Wolko (2004) deployed a RIL population between large- and small-seeded parents (Wt10245 × Wt11238) and reported five substantial QTLs for total protein content accounting for 18.3–25.5% of variation. Among these, three (prot-2, prot-3, and prot-4) were located in LG-V with high LOD values (4.4–5.3), whereas the two other QTLs, viz., prot-1 in LG-II and prot-5 in LG-VII revealed lower LOD (2.2 and 2.4). Afterwards, mapping populations were developed by crossing three pea cultivars with differential protein content, viz., Cameor, VavD265, and Ballet. In Cameor × VavD265, 7 LGs were detected with 6952 markers with a map length of 752.6 cM, whereas, in cross between VavD265 × Ballet, 850.1 cM map distance was covered with 6188 markers (Bourgeois et al. 2011; Tayeh et al. 2015b).

Globulins are the main kind of seed storage protein in peas which is further classified in to 11 S legumins and 7 S vicilins/convicilins. The distribution of these two classes of protein is influenced by genotype and the growing environments (Casey et al. 2001). Genetic analysis and DNA hybridization as well as sequence characterization of DNA and protein in pea detected four classes of legumin and five classes of 7 S protein genes and their respective QTLs (Newbigin et al. 1990). For fine characterization of globulin protein genes in pea, integrating two-dimensional electrophoresis (2D)-based proteomic and QTL mapping strategy was deployed involving 157 RIL population developed from three contrasting parents (Cameor, VavD265, and Ballet) with differential protein level by the Composite Interval Mapping (CIM) procedure. It was detected that 40 multigene families were involved in regulation of pea storage proteins where four gene classes were uncovered which are responsible for legumin biosynthesis in pea (Casey et al. 2001). The framework of three sets of legumin genes were well characterized by utilizing F_2 and F_6 individuals from selected crosses exploring RFLPs. They were mapped on chromosome 7 close to *r*, while the other maps to a locus near *a* on chromosome 1. The third class of legumin gene was likewise connected to *a*, according to the findings of one of the tested crosses (Domoney et al. 1986). The mRNA from developing pea seeds was used for construction of cDNA plasmid bank for characterization of vicilin coding genes, which further unveiled two different classes of vicilin genes that were initially synthesized as provicilin with subsequent processing at C-terminal peptide as well as post translational endo-proteolytic cleavage. The accumulation of convicilin was chiefly regulated by cis-regulatory regions whereas; both cis- and trans-regulatory regions were governing vicilins and legumins accumulation in pea. In vitro protein digestibility and protein composition appear to be significantly regulated by LG-IIa (Bourgeois et al. 2011). Recently, genome sequence information of the reference genome of cv. Cameor annotated 12 genes for legumin; 9 genes for vicilin and 2 genes for convicilin along with RY motifs (CATGCATG) associated with seed specific transcriptional regulation in pea (Kreplak et al., 2019). *LecA* gene was discovered which is the key gene governing lectin content in pea as well as associated with symbiotic N_2 fixation with *Rhizobium* in pea nodules (Domoney et al. 2013). Lectin is considered as an antinutritional factor due to poor digestibility. A knockout mutant was detected in pea with loss of function of *LecA* gene without any compensation regarding symbiotic N_2 fixation (Rayner et al. 2018). Similarly, *PA2* gene is governing another undesirable protein fraction albumin in pea seed.

Reference genome of cv. Cameor detected all together nine genes governing *PA2* and eight genes coding for another major albumin, *PA1* (Kreplak et al. 2019). However, in recent investigation it was detected that pea albumin having adequate sulphur containing amino acids which are essential for human being. Lipoxygenases (LOX), a subclass of seed storage proteins that controls the production of hydroperoxides from fatty acids, were found in pea. In pea seeds, there are two major LOX polypeptides (LOX-2 and LOX-3), and two or three genes at the *lox* locus positioned on LG IV encode these polypeptides (North et al. 1989). In recent years, attempt has been made to tag the genomic regions related with carbohydrate metabolism and seed storage protein accumulation in pea and to detect the SNP variation related with various HR genes (amylose, total starch, and crude protein concentration) for future marker assisted breeding programme (Jha et al. 2015b).

High density linkage map with less confidence interval is useful for QTL detection of major agronomic and economic traits in many crops including pea. GBS approach with simultaneous SNP detection has been extended for construction of high-resolution maps in pea to unveil the QTLs for seed mineral content (Ma et al. 2017). The study detected 6, 37, and 46 QTLs for seed weight, seed mineral content, and seed mineral concentration respectively, which explained phenotypic variation (PV) ranging from 2.4% to 43.3%. Another study by Gali et al. (2018) also deployed GBS approach for simultaneous detection of genome wide SNPs and QTLs associated with dietary fibre, seed protein, and mineral concentration using three RILs namely, PR-02, PR-07, PR-15 derived by cross of Orb × CDC Striker, Carrera × CDC Striker, and 1–2347-144 × CDC Meadow, respectively, at Crop Development Centre (CDC), University of Saskatchewan, Canada, employing single seed descent method. QTL mapping was executed employing CIM through QTL cartographer. Seed protein concentration QTLs have been detected on LG-Ib and LG-IVa of PR-02 with maximum LOD values of 5.0 and 3.4, respectively, and accounted up to 15.9 and 10.3% of the PV. Notably, QTLs were also detected for different fractions of dietary fibers. Four linkage groups were found for acid detergent fibre (ADF), including two QTLs on LG VII that individually accounted for 28.0% and 26.2% of the PV in the PR-02 population. Significant QTLs for ADF were discovered on LG IV and VIIa for the PR-07 population. With regard to neutral detergent fibre (NDF), in the PR-02 population, identified QTLs on LG-Va, and in the PR-07 RIL population, identified QTLs on LG Ia, IV, and VIIa, accounting for up to 44% of phenotypic variance. On LG-IIb and LG-IVa of PR-02, LG-Ia, LG-IIIb, LG-IIIc, and LG-VIIb of PR-07, numerous QTLs for seed starch content were found. The largest LOD value represented by LG-VIIa was 8.4, and 20.1% of the phenotypic variance was elucidated by this QTL. QTLs for seed iron (Fe) concentration were found over four LGs of PR-02, and the QTL on LG-IIIb was highly significant with LOD values of 7.6 and 6.3. Based on two phenotypic experiments, three linkage groups of the PR-02 population were shown to have QTLs for seed selenium (Se) content. Of these, LG-VII embodied QTLs within the same linkage group region in both trials and demonstrated a phenotypic variance up to 15.0%. In two of the six trials for seed Se content in the PR-7 population, QTLs on LG-IV and LG-Vb were found. For seed zinc (Zn) concentration, four QTLs were found: two on LG-VI of PR-02 and one

each on LG-Ia and LG-IIIb. With a LOD value of 13.7 and a PV of 25.8%, the QTL on LG-IIIb was highly significant. QTLs for seed phytate content were mapped in PR-15 population. Three linkage groups, viz., LG-IIIa, LG-V, and LG-Via were shown to include QTLs for this characteristic, with the LG-V QTL being found in all four trials evaluated in two sites (Gali et al. 2018). The recognized QTLs linked with HR genes in pea are valuable treasure for marker assisted breeding programme for improving seed quality traits in pea.

7 Map-Based Cloning of HR Genes/QTLs

An innovative method that pinpoints the underlying origin of a genetic variation is map-based cloning (MBC). Without prior knowledge of specific genes, MBC can access a practically infinite resource of induced and natural genetic variation. Therefore, scientists are attempting to use this method in model plants to clone the orthologues genes in related plant species. *Medicago truncatula* and *Lotus japonicus* are considered as model legumes as these are diploid with eight and six chromosomes, respectively, and having relatively small genome size of around 500 Mb (Cannon et al. 2009). These model legume species facilitate to transfer the genomic resources into target legume crops for better understanding of development as well as evolutionary biology. High degree of synteny between model species and pea facilitate to deploy cross species gene-based markers for unveiling the homologous genome segments. A comparison of the genomes of pea (*P. sativum*) and alfalfa (*M. truncatula*) has been done through comparing linkage maps of both the species to understand the co-linearity between the linkage groups and the presence of conserved and orthologues genes using gene-based RFLP and PCR markers (Kalo et al. 2004). Considering the model legume *M. truncatula,* it was discovered that the *ABI5* transcription factor is a key factor in determining the accumulation of seed globulins. This study also characterized loss-of-function *abi5* mutants in pea, showing that the mutants had higher levels of other important seed proteins and lower vicilin concentrations (Le Signor et al. 2017). Further, SNP variation was detected in *O2like* gene in pea that is similar to *opaque2* in maize, a bZIP transcription regulating factor with having significant role in management of protein content (Jha et al. 2015b).

Starch or amylum is the most abundant form of storage carbohydrate in pea seeds composed of amylose and highly branched amylopectin. Attempt has been made to characterize the key genes for getting better insight about the starch biosynthetic pathways through generating array of mutants with differential starch concentration. Various mutant forms (*Iam*, *r*, *rb*, *rug3*, *rug4,* and *rug5*) have been generated targeting the key enzymes like granule bound starch synthase Ia (GBSSIa) (Dry et al. 1992), starch branching enzyme I (SBEI) (Burton et al. 1995;), ADP-glucose pyrophosphorylase (AGPase) (Burgess et al. 1997), phosphoglucomutase [Plastidial] (PGM) (Harrison et al. 1998), sucrose synthase (SuSy1) (Barratt et al. 2001), and starch synthase II (SSII) (Craig et al. 1998). The generated mutant in pea with altered or absence of the targeted genes also exhibited pleotropic effects on the seed shape

changed from round to wrinkled with a 50% reduction of starch content and an increase in lipid and sucrose content.

Large and stable genomic clones for crops with complex genome are frequently created and stored using bacterial artificial chromosomes (BACs), which are employed as a genomic library. The first BAC library in pea consisting of partially HindIII-digested DNA fragments with a mean size of 105 kb was developed considering pea cv. PI269818 for isolation of disease resistance genes and genes governing economic traits (Coyne et al. 2007). A sequence-based physical mapping technique called whole genome profiling (WGP) makes use of sequence tags produced by NGS. A recent study by Gali et al. (2019) utilized pea cv. 'Cameor' for construction of BAC library for localization of the gene through physical mapping for further structural and functional annotation followed by genome sequencing as well as positional cloning of the mined genes having economic importance.

8 Omics Approaches in Relation to HR Genes

In recent decades, the newly emerging omics-based approaches facilitate towards locating genes/QTLs governing the traits of interest and further make available the robust markers platform for marker assisted breeding programme. Microarray or RNA sequencing technology as well as the recent cutting edge NGS based transcriptome assembly can generate valuable genomic resources and markers that can be assembled for construction of highly saturated genome maps, improve genotyping technologies (sequencing and resequencing), and identification of additional markers like SNPs and EST-SSRs (Kaur et al. 2012). Pea had relatively meagre genomic resources compared to the model legume until the first pea reference genome assembly (Kreplak et al. 2019) became available. This offered pea genomics-assisted breeding a new momentum. The estimated pea genome is 3.92 Gb in size, spanning 88% of the reported genome assembly. This reference genome assembly has offered several contigs, transcripts, markers, and GBS platforms for their further annotation. The creation of integrated linkage maps and the comparison of markers across various studies have been made easier by the alignment of sequence reads to a common reference genome sequence. In pea, the transcriptome analysis has been mostly carried out to unveil the host pathogen interaction and pathogenesis related proteins (Winter et al. 2016), genes, and transcripts related with the mechanism underlying abiotic stress resistance (Chen et al. 2013; Bahrman et al. 2019), and there is scantiness regarding transcriptome analysis in relation to HR genes. However, the gene expression atlas developed in pea using the variety "Little Marvel" (Franssen et al. 2011) or by using another two cultivars (Kaspa and Parafeld) by Sudheesh et al. (2015) and Alves-Carvalho et al. (2015) generated valuable transcripts atlas that will further boost up reverse genetic studies, fine mapping, allele mining, and identification of candidate genes. In another study, eight diverse accessions of pea and five RIL populations were deployed to generate genome wide transcriptome-based SNP arrays using Illumina

Golden Gate assay (Sindhu et al. 2014). A total of 1536 polymorphic SNP loci were found, and by combining them with previously discovered anchor markers, the first high-density pea SNP map was created, delineating all seven linkage groups. This breakthrough can expedite tagging and mapping the QTLs related with agronomic and HR genes for boosting up seed quality for betterment of human nutrition. Using 12 RIL populations, GenoPea 13.2 K SNP Array has been developed in pea with construction of collinear map of 3918–8503 SNPs and a total of 12,802 transcript-derived SNP markers were available for future genomic study (Tayeh et al. 2015a). The high resolution and high-density consensus map developed with 15,079-marker will further facilitate the identification of ohnologue-rich regions within the pea genome thus strengthen pea genomic resources.

Proteomics and Metabolomics are new emerging arena to study the function of novel proteins associated with various physiological and biological events. Proteomics defined as the high-throughput study of proteins has assumed a leading role in plant biological research and stress responses owing to availability of plant genome sequence information in public database. Moreover, with the advancement of various high throughput approaches like mass spectrometry (MS), Tandem mass spectrometry and other quantitative assay coupled with bioinformatics approaches have offered easy characterization, quantification, and further validation of array of functional proteins from any crop species (Ramalingam et al. 2015). In addition to proteomics, metabolomics is a crucial tool of functional genomics that links cellular metabolic activity with phenotypes through the identification and quantification of metabolomes, a collection of metabolites or small molecules, within a cell, tissue, or organism (Weckwerth 2003). Like transcriptome assembly in pea, most of the proteomics and metabolomics study has been carried out in relation to characterization of proteins and metabolites underlying stress responses (Ranjbar Sistani et al. 2017) with meagre information that can uncover novel HR proteins and metabolites in pea. Improving vicilin concentration in pea is the need of the hour as vicilin having lower concentration of sulphur containing amino acid in comparison to legumin. Proteomic study by Bourgeois et al. (2009) was the first pea mature seed proteome reference map, detected 156 novel proteins including 24 different genes controlling vicilin in pea. This study provided the diversity in relation to seed storage proteins and their plasticity during plant developmental stages. Later on, integrating proteomic and QTL mapping approaches, Bourgeois et al. (2011) developed proteomic atlas from the RIL population developed from three pea genotypes Cameor, VavD265 and Ballet and unveiled the genetic architecture regarding the seed proteome variability. Protein quantity loci (PQL) were hunted for 525 spots noticed on 2-D gels and interestingly, most PQL were mapped in clusters. This demonstrated how a few numbers of loci were responsible for accumulating the major store protein families. Previous study also reported how fast-neutron mutagenesis may be used to disrupt many genes at a single locus, affecting the accumulation of vicilin and other proteins (Domoney et al. 2013). It has been demonstrated that a mutant allele at the Vc-2 gene affects seed nitrogen contents and major vicilin polypeptide synthesis (Chinoy et al. 2011).

9 Genomics-Aided Breeding for HR Traits

Seed quality traits including protein, carbohydrate, and mineral concentration are mostly complex and quantitative in nature. Multiple genes, environment, and genotype × environment interaction create ambiguities towards expression of these HR genes related with seed quality. Besides biparental linkage mapping, genome wide association mapping (GWAS) is nowadays used to understand these complex traits. Association mapping based on the principle of linkage disequilibrium can capture greater number of alleles originated during the course of evolution. The first GWAS was applied to understand the allelic variation regarding starch metabolic pathway that can affect the chain length distribution (CLD) and starch structure in a set of 92 diverse pea accessions (Carpenter et al. 2017). Associations for polymorphisms in seven potential genes and the Mendel's r locus were discovered that govern round versus wrinkled seed phenotype. Amongst seven significant candidate genes, three genes, viz., *r* (rugosus allele), *UGPase* (UDP-glucose pyrophosphorylase), and *AGPS2* (ADP-glucose pyrophosphorylase S2 subunit) portrayed a significant relationship with CLD, and the amylose content was linked with the *r* locus. Another GWAS was performed to track the loci related with agronomic, seed morphology, and seed quality traits (protein, carbohydrate, and fiber content) using 135 diverse accessions of pea from 23 different pea growing countries. GBS approach was also integrated for detection of 16,877 high quality SNP arrays. Five SNPs were recognized that were being linked with ADF concentration (positioned on chromosomes 5, 6, and 7), whereas, eight SNPs were detected with NDF content (chromosomes 2, 3, 5, 6 and 7). Four markers (Chr1LG6_176606388, Chr2LG1_457185, Chr3LG5_234519042, and Chr7LG7_8229439), located on chromosomes1, 2, 3, and 7, were connected with seed starch content, and one marker (Chr3LG5_194530376) identified was linked with seed protein concentration (Gali et al. 2019). Dissanayaka et al. (2020) used a panel of 135 diverse pea accessions for identification of trait associated SNP arrays for an association study concerning seed mineral content. Out of the 16,877 SNP markers deployed for association analysis, five each were recognized for relationship with Fe and Zn content in pea seeds. Markers detected in the present study related with Fe (Chr5LG3_204123886) and Zn (Chr5LG3_1921113554, Chr5LG3_197808492, and Sc4026_15361) would be useful for MAS programme in pea for rapid generation of cultivar with good seed quality traits.

With the expansion of pea genomic resources in recent years, particularly the reference genome sequence, which makes it easier to comprehend the allelic variation underlying the important traits and facilitate towards development of better pea cultivars opting various genomics assisted breeding programme like haplotype-based breeding, genome editing, and genomic selection (GS). Plant breeders are quickly adopting GS, a potent breeding technique, especially for traits that are challenging to detect. In the next 10 years, breeding for increased pea productivity and quality is anticipated to benefit from the use of GS in conjunction with high-throughput SNP genotyping platform. In order to maximize genomic prediction for complex traits, GS models depend on training and testing populations and enumerate the Genomic Estimated Breeding Values (GEBV). Comparatively to conventional

marker-assisted selection, GS enables the simultaneous selection of several attributes. GS programme initiated in pea with the easily observable agronomic traits characterized by high heritability like seed weight, days to flowering, pods per plant, Ascochyta blight resistance, etc., with prediction accuracy of 0.19–0.84% (Tayeh et al. 2015b). These research findings offer "proof of concept" for the idea that selecting pea breeding lines for HR traits in various environments can be made more effective in the future by combining superior training and test population sets and integrating robust and reliable prediction models and high-density marker system.

10 Other Approaches and Future Strategies for Modulating HR Genes in Pea

Like other legume, pea seeds are also fantastic treasury of dietary protein. Although they consist of numerous protein classes that resist proteolysis to varying grades and are adversely correlated with quality and animal health. Protease inhibitors, specifically trypsin/chymotrypsin inhibitors (TI), present in the pea seeds are regarded as antinutritional components that usually require additional heat-treatment or other processing for use as feed for cattle and poultry (Patto et al. 2015). Pea seed TI are mostly of the Bowman-Birk inhibitor (BBI) category, with wide genetic variation exists regarding inhibitory activity (Domoney and Welham 1992). Two closely related genes, *TI1* and *TI2*, have been found to encode the two primary BBI isoforms that are expressed in pea seeds and mostly inhibit both trypsin and chymotrypsin (Domoney et al. 2002). Exploitation of a Targeted Induced Local Lesions IN Genomes (TILLING) population in pea facilitated to detect novel genetic variants regarding various classes of inhibitor proteins, which can provide deeper insight regarding the structure-function and relationships within the protease inhibitors. Moreover, targeted screening within *TI1* gene of pea enabled to detect natural variant devoid of inhibitory activity class of seed protein. A total of 13 nucleotide changes have been detected including seven changes were in noncoding regions and six alterations influencing the coding sequence that generated missense mutations. It was validated from the finding that the substitution of C77Y in the mature mutant inhibitor completely eliminated inhibitor activity. A *P. elatius* accession as a double null mutant for the two closely linked genes *TI1* and *TI2* was detected with extremely low seed protease inhibitory activity, and introgression of the mutant into cultivated germplasm has been achieved (Clemente et al. 2015).

In pea, attempt has been made to introduce antisense construct for genetic manipulation of the inhibitor genes. The promoter from *TI* gene has been isolated, characterized, and reintroduced within pea by *Agrobacterium*-mediated genetic transformation, as a TI promoter-bglucuronidase (GUS) gene fusion. A second gene construct uses the *TI1* gene promoter for direct exhibition of an antisense *TI* gene. Seed *TI* activities in some transformed pea lines with this construct were tested, and it was detected that the activity of inhibitor genes was reduced significantly (Welham and Domoney 2000). Attempt was also being made to understand the structural relationship of multiple genes responsible for governing legumin

biosynthesis in developing pea seeds by Casey et al. (2001). Earlier report confirmed that in pea seeds legumin biosynthesis was usually modified due to the nature of the *r* loci governing the structural gene for starch-branching enzyme isoform I (*SBE I*) or *rb* responsible for synthesis of ADP-glucose pyrophosphorylase. This ultimately led to changes in the ratio of legumin: vicilin within the total seed protein in pea. It was detected that the double mutants (*rrb*) showed a significant drop in the quantity of legumin without altering the concentration of vicilin. Further expression of cloned legumin cDNA construct of pea in transgenic wheat seeds able to synthesize an array of paracrystalline legumins. This evidence validated that, in spite of the presence of multiple genes with structural heterogeneity, the pea legumin comprises of a single type of subunit.

In pea, there is no successful report of genome editing due to the calcitrant nature of this legume. However, the most recent approaches, such as de novo meristem induction by and DNA transfer in mature plants enabled by nanomaterial offered great promise towards the development of genome-edited pea plants with advantageous features.

With the advent of affordable genomic tools, access to pea genomic resources will increase in the near future. This will help to close knowledge gaps regarding allelic variation and the genetic mechanisms governing the metabolism of various HR genes and will hasten the creation of novel pea cultivars through genomic assisted breeding programme with refine HR genes for the benefit of mankind.

11 Future Prospects and Conclusion

Malnutrition has become a serious problem owing to the insufficient supply of nutritionally balanced diets to the resource-poor folks of under developed/developing countries in the prevailing changing climate. Unfortunately, hidden hunger is steadily increasing due to the reliance of the increasing population on cereal based carbohydrate rich diet which is deprived of vitamins and essential micronutrients. Various strategies such as dietary supplementation, fortification of foods, and agrofortification are being used to upsurge the disposal of an essential nutrient in the daily diet, but these are not accessible to the resource-poor population. Thus, genetic biofortification of edible crops through conventional and modern breeding approaches is considered as most effective, economic, and sustainable approach for the entire stakeholders. In the case of peas, very limited systematic efforts have been made to assess the extent of variability for nutritionally important traits such as protein, iron, zinc, folate, vitamins, phenolic compound, and selenium. In addition, the variability so far observed could not be exploited judiciously in the regular breeding programme toward genetic biofortification. Therefore, it is the need of the hour that large-scale high-throughput phenotyping should be done to unveil the genetic variability that exists in peas. Unfortunately, unlike other legume crops, restricted efforts have been made to explore the magnitude of variability that exists

for nutritionally important traits in landraces and crop wild relatives. Thus, proper phenotyping of landraces and CWR needs to be done and subsequently, the identified sources must be embraced in a regular breeding programme to develop nutrient-enriched pea. Indeed, conventional breeding approaches have been well accepted by the public as being cost-effective, free of synthetic inputs, and eco-friendly and sustainable practices, but their laborious and time-consuming nature prompted plant breeders to opt for genomics-assisted breeding. Genomic resources facilitate breeders in the exploitation of existing genetic variability more precisely with minimum duration and expenses. Unlike other legume crops, in pea, the genomic resources have not been developed and exploited to identify genes/QTL for nutritional traits. Therefore, the recently emerged omics-based approaches need to be embraced to facilitate the identification of genes/QTLs and further make available the robust markers platform for a marker-assisted breeding programme in pea. Microarray or RNA sequencing technology and NGS-based transcriptome assembly could be used to generate valuable genomic resources and markers that can be assembled for the construction of highly saturated genome maps, improve genotyping technologies (sequencing and resequencing), and identification of additional markers like SNPs and EST-SSRs. Notably, the pea reference genome assembly offered several contigs, transcripts, markers, and GBS platforms for their further annotation. The transcriptomics, proteomics, and metabolomics analysis have been mostly carried out in pea to unveil the mechanism underlying abiotic and biotic stress resistance, and there is insufficiency regarding analysis of nutrition-related genes, which must be accelerated. However, most recently few valuable transcripts atlas has been generated that will further boost reverse genetic studies, fine mapping, allele mining, and identification of candidate genes. In case genetic variability is inadequate for traits of interest then the desired variability may be created by adopting induced mutagenesis or modern techniques like genome editing (i.e., clustered regularly interspaced short palindromic repeats (CRISPR)- associated system (CRISPR/Cas). However, to date, there is no successful report of genome editing due to the recalcitrant nature of this legume. In addition, exploitation of a Targeted Induced Local Lesions IN Genomes (TILLING) population in peas would facilitate the detection of novel genetic variants regarding various classes of inhibitor proteins which can provide deeper insight into the structure-function and relationships within the protease inhibitors. Another powerful tool is a genetic transformation or transgenic, which possess many advantages including its long-term sustainability. In addition to protein, peas also consist of numerous antinutritional components specifically trypsin/chymotrypsin inhibitors (TI) and phytic acid. Therefore, gene silencing approaches should be used to down-regulate genes encoding antinutritional compounds without affecting other pathways and processes involved in plant growth and development. The advent of affordable genomic tools and access to reference genome will help to reduce the knowledge gaps regarding allelic variation and the genetic mechanisms governing the metabolism of various nutrition-related genes and will hasten the creation of novel pea cultivars with refined HR genes for the benefit of mankind.

References

Alves-Carvalho S, Aubert G, Carrère S et al (2015) Full-length denovo assembly of RNA-seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights in to root nodulation in this species. Plant J 84:1–19. https://doi.org/10.1111/tpj.12967

Antonets KS, Belousov MV, Sulatskaya AI et al (2020) Accumulation of storage proteins in plant seeds is mediated by amyloid formation. PLoS Biol 18:e3000564. https://doi.org/10.1371/journal.pbio.3000564

Aubert G, Morin J, Jacquin F et al (2006) Functional mapping in pea, as an aid to the candidate gene selection and for investigating synteny with the model legume *Medicago truncatula*. Theor Appl Genet 112(6):1024–1041

Bahrman N, Hascoët E, Jaminon O et al (2019) Identification of genes differentially expressed in response to cold in *Pisum sativum* using RNA sequencing analyses. Plan Theory 8:288

Barratt DP, Barber L, Kruger NJ et al (2001) Multiple, distinct isoforms of sucrose synthase in pea. Plant Physiol 127(2):655–664

Bateson W (1901) Experiments in plant hybridization by Gregor Mendel. J Royal Hortic Soc 24(I):1–32

Bhattacharyya MK, Smith AM, Ellis TN et al (1990) The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. Cell 60(1):115–122

Bhattacharyya MK, Martin C, Smith A (1993) The importance of starch biosynthesis in the wrinkled seed shape character of peas studied by Mendel. Plant Mol Biol 22:525–531. https://doi.org/10.1007/BF00015981

Bishnoi S, Khetarpaul N, Yadav RK (1994) Effect of domestic processing and cooking methods on phytic acid and polyphenol contents of pea cultivars (*Pisum sativum*). Plant Foods Hum Nutr 45: 381–388

Blair MW, Astudillo C, Rengifo J et al (2011) QTL for seed iron and zinc concentrations in a recombinant inbred line population of Andean common beans (*Phaseolus vulgaris* L.). Theor Appl Genet 122:511–521

Bourgeois M, Jacquin F, Savois V et al (2009) Dissecting the proteome of pea mature seeds reveals the phenotypic plasticity of seed protein composition. Proteomics 9:254–271. https://doi.org/10.1002/pmic.200700903

Bourgeois M, Jacquin F, Cassecuelle F et al (2011) A PQL (protein quantity loci) analysis of mature pea seed proteins identifies loci determining seed protein composition. Proteomics 11:1581–1594. https://doi.org/10.1002/pmic.201000687

Burgess D, Penton A, Dunsmuir P et al (1997) Molecular cloning and characterization of ADP-glucose pyrophosphorylase cDNA clones isolated from pea cotyledons. Plant Mol Biol 33(3):431–444

Burstin J, Marget P, Huart M et al (2007) Developmental genes have pleiotropic effects on plant morphology and source capacity, eventually impacting on seed protein content and productivity in pea. Plant Physiol 144:768–781. https://doi.org/10.1104/pp.107.096966

Burton RA, Bewley JD, Smith AM et al (1995) Starch branching enzymes belonging to distinct enzyme families are differentially expressed during pea embryo development. Plant J 7(1):3–15

Cannon SB, May GD, Jackson SA (2009) Three sequenced legume genomes and many crop species: rich opportunities for translational genomics. Plant Physiol 151(3):970–977

Carpenter MA, Shaw M, Cooper RD et al (2017) Association mapping of starch chain length distribution and amylose content in pea (*Pisum sativum* L.) using carbohydrate metabolism candidate genes. BMC Plant Biol 17(1):1–23. https://doi.org/10.1186/s12870-017-1080-9

Casey R, Davies DR (1993) Peas: genetics, molecular biology and biotechnology. CAB International, Wallingford, 323 p

Casey R, Christou P, Domoney C et al (2001) Expression of legumin and vicilin genes in pea mutants and the production of legumin in transgenic plants. Food Nahrung 45(6):385–387

Chen H, Osuna D, Colville L et al (2013) Transcriptome-wide mapping of pea seed ageing reveals a pivotal role for genes related to oxidative stress and programmed cell death. PLoS One 8:e78471
Cheng P, Holdsworth W, Ma Y et al (2015) Association mapping of agronomic and quality traits in USDA pea single-plant collection. Mol Breed 35(2):1–13
Chinoy C, Welham T, Turner L et al (2011) The genetic control of seed quality traits: effects of allelic variation at the Tri and Vc-2 genetic loci in *Pisum sativum* L. Euphytica 180(1):107–122
Clemente A, Arques MC, Dalmais M et al (2015) Eliminating antinutritional plant food proteins: the case of seed protease inhibitors in pea. PLoS One 10:e0134634
Coyne CJ, McClendon MT, Walling JG et al (2007) Construction and characterization of two bacterial artificial chromosome libraries of pea (*Pisum sativum* L.) for the isolation of economically important genes. Genome 50:871–875. https://doi.org/10.1139/G07-063
Craig J, Lloyd JR, Tomlinson K et al (1998) Mutations in the gene encoding starch synthase II profoundly alter amylopectin structure in pea embryos. Plant Cell 10:413–426. https://doi.org/10.1016/S1369-5266(98)80075-7
Dahl WJ, Foster LM, Tyler RT (2012) Review of the health benefits of peas (*Pisum sativum* L.). Br J Nutr 108(S1):S3–S10
Demirbas BA (2018) Micro and macronutrients diversity in Turkish pea (*Pisum sativum*) germplasm. Int J Agric Biol 20(4):701–710
Diapari M, Sindhu A, Warkentin TD et al (2015) Population structure and market trait association studies of iron, zinc and selenium concentrations in seed of field pea (*Pisum sativum* L.). Mol Breed 35:30–42. https://doi.org/10.1007/s11032-015-0252-2
Díaz CL, Melchers LS, Hooykaas PJJ et al (1989) Root lectin as a determinant of host-plant specificity in the *Rhizobium*-legume symbiosis. Nat 338:579. https://doi.org/10.1038/338579a0
Dissanayaka DN, Gali KK, Jha AB et al (2020) Genome-wide association study to identify single nucleotide polymorphisms associated with Fe, Zn, and Se concentration in field pea. Crop Sci 60 (4):2070–2084
Dixit GP, Parihar AK, Gupta S (2014) Perspective for increasing fieldpea production in India. In: Handbook on minor and imported pulses of India. commodityindia.com, pp 37–39
Domoney C, Welham T (1992) Trypsin inhibitors in *Pisum*: variation in amount and pattern of accumulation in developing seed. Seed Sci Res 2(3):147–154
Domoney C, Ellis THN, Davies DR (1986) Organization and mapping of legumin genes in *Pisum*. Mol Gen Genet 202:280–285. https://doi.org/10.1007/BF00331651
Domoney C, Welham T, Ellis N et al (2002) Three classes of proteinase inhibitor gene have distinct but overlapping patterns of expression in *Pisum sativum* plants. Plant Mol Biol 48:319–329. https://doi.org/10.1023/A:1013379430582
Domoney C, Knox M, Moreau C et al (2013) Exploiting a fast neutron mutant genetic resource in *Pisum sativum* (pea) for functional genomics. Funct Plant Biol 40:1261. https://doi.org/10.1071/FP13147
Dry I, Smith A, Edwards A et al (1992) Characterization of cDNAs encoding two isoforms of granule-bound starch synthase which show differential expression in developing storage organs of pea and potato. Plant J 2(2):193–202
Duenas M, Estrella I, Hernandez T (2004) Occurrence of phenolic compounds in the seed coat and the cotyledon of peas (*Pisum sativum* L.). Eur Food Res Technol 219:116–123
Ekvall J, Stegmark R, Nyman M (2006) Content of low molecular weight carbohydrates in vining peas (*Pisum sativum*) related to harvest time, size and brine grade. Food Chem 94:513–519
Ellis N, Hattori C, Cheema J et al (2018) NMR metabolomics defining genetic variation in pea seed metabolites. Front Plant Sci 9:1022
Fairweather-Tait SJ, Bao Y, Broadley MR et al (2011) Selenium in human health and disease. Antioxid Redox Signal 14:1337–1383. https://doi.org/10.1089/ars.2010.3275
Fernando WMU, Hill JE, Zello GA et al (2010) Diets supplemented with chickpea or its main oligosaccharide component raffinose modify fecal microbial composition in healthy adults. Benefic Microbes 1:197–207

Forster C, North H, Afzal N et al (1999) Molecular analysis of a null mutant for pea (*Pisum sativum* L.) seed lipoxygenase-2. Plant Mol Biol 39(6):1209–1220

Franssen SU, Shrestha RP, Bräutigam A et al (2011) Comprehensive transcriptome analysis of the highly complex Pisum sativum genome using next generation sequencing. BMC Genomics 12: 227. https://doi.org/10.1186/1471-2164-12-227

Gali KK, Liu Y, Sindhu A et al (2018) Construction of high-density linkage maps for mapping quantitative trait loci for multiple traits in feld pea (*Pisum sativum* L.). BMC Plant Biol 18:172

Gali KK, Sackville A, Tafesse EG et al (2019) Genome-wide association mapping for agronomic and seed quality traits of field pea (*Pisum sativum* L.). Front Plant Sci 10:1538. https://doi.org/10.3389/fpls.2019.01538

Gaur V, Qureshi IA, Singh A et al (2010) Crystal structure and functional insights of hemopexin fold protein from grass pea. Plant Physiol 152:1842–1850. https://doi.org/10.1104/pp.109.150680

Gawłowska M, Święcicki W, Lahuta L et al (2017) Raffinose family oligosaccharides in seeds of *Pisum* wild taxa, type lines for seed genes, domesticated and advanced breeding materials. Genet Resour Crop Evol 64(3):569–578

Ge J, Sun CX, Corke H et al (2020) The health benefits, functional properties, modifications, and applications of pea (*Pisum sativum* L.) protein: current status, challenges, and perspectives. Compr Rev Food Sci Food Saf 19(4):1835–1876

Gebreegziabher BG, Tsegay BA (2020) Proximate and mineral composition of Ethiopian pea (*Pisum sativum* var. *abyssinicum* A. Braun) landraces vary across altitudinal ecosystems. Cogent Food Agri 6(1):1789421

Grusak MA, Cakmak I (2005) Methods to improve the crop-delivery of minerals to humans and livestock. In: Broadley MR, White PJ (eds) Plant nutritional genomics. Blackwell Publishing, Oxford, pp 265–286

Guillon F, Champ MJ (2002) Carbohydrate fractions of legumes: uses in human nutrition and potential for health. Br J Nutr 88(S3):293–306

Gupta S, Parihar AK (2015) Fieldpea cultivation in India (Pocket guide). AICRP on MULLaRP, Indian Institute of Pulses Research, Extension Bulletin, Kanpur. 35pp

Han JY, Tyler RT (2003) Determination of folate concentrations in pulses by a microbiological method employing trienzyme extraction. J Agric Food Chem 51:5315–5318

Harmankaya M, Özcan MM, Karadaş S et al (2010) Protein and mineral contents of pea (*Pisum sativum* L.) genotypes grown in Central Anatolian region of Turkey. South-West J Hortic Biol Environ 1(2):159–165

Harrison CJ, Hedley CL, Wang TL (1998) Evidence that the rug3 locus of pea (*Pisum sativum* L.) encodes plastidial phosphoglucomutase confirms that the imported substrate for starch synthesis in pea amyloplasts is glucose-6-phosphate. Plant J 13:753–762

He X, Chen J, He X et al (2020) Industryscale microfluidization as a potential technique to improve solubility and modify structure of pea protein. Innov Food Sci Emer Technol 67:102582

Hotz C, Brown KH (2004) Assessment of the risk of zinc deficiency in populations and options for its control. Food Nutr Bull 25:S94–S204

Hussain H, Martin C (2009) Comparative analysis of primary and secondary structure for pea isoamylase isoforms predicts different catalytic properties against glucan substrates. Starch-Stärke 61(10):570–577

Irzykowska L, Wolko B (2004) Interval mapping of QTLs controlling yield-related traits and seed protein content in Pisum sativum. J Appl Genet 45:297–306

Jha AB, Ashokkumar K, Diapari M et al (2015a) Genetic diversity of folate profiles in seeds of common bean, lentil, chickpea and pea. J Food Compos Anal 42:134–140

Jha AB, Taran B, Diapari M et al (2015b) SNP variation within genes associated with amylose, total starch and crude protein concentration in field pea. Euphytica 206:459–471. https://doi.org/10.1007/s10681-015-1510-4

Kalo P, Seres A, Taylor SA et al (2004) Comparative mapping between *Medicago sativa* and *Pisum sativum*. Mol Genet Genomics 272:235–246

Kaur S, Pembleton LW, Cogan NO et al (2012) Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. BMC Genomics 13:104. https://doi.org/10.1186/1471-2164-13-104

Khamparia RS, Singh MV, Sharma BL et al (2010) AICRP: micro and secondary nutrients and pollutant elements in soil and plants. Indian Inst Soil Sci Bhopal Madhya Pradesh 6:1–113
Kooistra E (1962) On the differences between smooth and three types of wrinkled peas. Euphytica 77:357–373
Krajewski P, Bocianowski J, Gawłowska M et al (2012) QTL for yield components and protein content: a multienvironment study of two pea (*Pisum sativum* L.) populations. Euphytica 183 (3):323–336
Kreplak J, Madoui M, C'apal P et al (2019) A reference genome for pea provides insight into legume genome evolution. Nat Genet 51:1411–1422. https://doi.org/10.1038/s41588-019-0480-1
Lamichaney A, Parihar AK, Dixit GP, Singh AK, Kumar N, Singh NP (2021a) Intact funiculus in mature harvested seeds of field pea (*Pisum sativum* L.): preliminary investigation and possible implications. Crop Sci 61(5):2863–2871
Lamichaney A, Parihar AK, Hazra KK et al (2021b) Untangling the influence of heat stress on crop phenology, seed set, seed weight, and germination in feld pea (*Pisum sativum* L.). Front Plant Sci 12:437
Le Gall M, Quillien L, S'eve B et al (2007) Weaned piglets display low gastrointestinal digestion of pea (*Pisum sativum* L.) lectin and pea albumin 2. J Anim Sci 85:2972–2981. https://doi.org/10.2527/jas.2006-795
Le Signor C, Aim'e D, Bordat A et al (2017) Genome-wide association studies with proteomics data reveal genes important for synthesis, transport and packaging of globulins in legume seeds. New Phytol 214:1597–1613. https://doi.org/10.1111/nph.14500
Li BW, Andrews KW, Pehrsson PR (2002). Individual sugars, soluble, and insoluble dietary fiber contents of 70 high consumption foods. J Food Compos Anal 15(6):715–723
Lockyer S, Nugent AP (2017) Health effects of resistant starch. Nutr Bull 42:10–41. https://doi.org/10.1111/nbu.12244
Lu ZX, He JF, Zhang YC et al (2020) Composition, physicochemical properties of pea protein and its application in functional foods. Crit Rev Food Sci Nutr 60(15):2593–2605
Ma Z, Boye JI, Simpson BK et al (2011) Thermal processing effects on the functional properties and microstructure of lentil, chickpea, and pea fours. Food Res Int 44(8):2534–2544
Ma Z, Boye JI, Azarnia S et al (2015) Volatile favour profle of Saskatchewan grown pulses as affected by different thermal processing treatments. Int J Food Prop 19(10):1–21
Ma Y, Coyne CJ, Grusak MA et al (2017) Genome-wide SNP identification, linkage map construction and QTL mapping for seed mineral concentrations and contents in pea (*Pisum sativum* L.). BMC Plant Biol 17(1):43
Macas J, Neumann P, Navrátilová A (2007) Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. BMC Genomics 8(1):1–16
Madoui MA, Labadie K, Aury JM et al (2015) The international pea genome sequencing project: sequencing and assembly progresses. In: Plant and animal genome XXIII conference, SanDiego, CA
McAdam EL, Meitzel T, Quittenden LJ et al (2017) Evidence that auxin is required for normal seed size and starch synthesis in pea. New Phytol 216:193–204
Meitzel T, Radchuk R, McAdam EL et al (2021) Trehalose 6-phosphate promotes seed filling by activating auxin biosynthesis. New Phytol 229:1553–1565
Mertens C, Dehon L, Bourgeois A et al (2012) Agronomical factors influencing the legumin/vicilin ratio in pea (*Pisum sativum* L.) seeds. J Sci Food Agric 92:1591–1596. https://doi.org/10.1002/jsfa.4738
Newbigin EJ, DeLumen BO, Chandler PM et al (1990) Pea convicilin: structure and primary sequence of the protein and expression of a gene in the seeds of transgenic tobacco. Planta 180:461–470. https://doi.org/10.1007/BF02411442
North H, Casey R, Domoney C (1989) Inheritance and mapping of seed lipoxygenase polypeptides in *Pisum*. Theor Appl Genet 77:805–808. https://doi.org/10.1007/BF00268330
Pandey S, Gritton ET (1975) Genotypic and phenotypic variances and correlations in peas. Crop SeL 15:353–356

Parihar AK, Dixit GP, Pathak V et al (2014a) Assessment of the genetic components and trait associations in diverse set of fieldpea (*Pisum sativum* L.) genotypes. Bang J Bot 43(3):323–330
Parihar AK, Dixit GP, Pathak V et al (2014b) Genetic diversity and trait inter-relationship studies in a diverse set of fieldpea (*Pisum sativum* L.) genotypes. J Food Legumes 27(4):297–301
Parihar AK, Bohra A, Dixit GP (2016) Nutritional benefits of winter pulses with special emphasis on peas and rajmash. In: Singh U, Praharaj CS, Singh SS (eds) Biofortifications of food crops. Springer publication, pp 61–71. https://doi.org/10.1007/978-81-322-2716-8_6
Parihar AK, Dixit GP, Bohra A et al (2020) Genetic advancement in dry pea (*Pisum sativum* L.): retrospect and prospect. In: Gosal SS, Wani SH (eds) Accelerated plant breeding, vol 3. Springer, Cham, pp 283–341
Parihar AK, Dixit GP, Singh U et al (2021) Potential of field pea as a nutritionally rich food legume crop. In: Gupta S, Kumar J (eds) Sen Gupta D. Breeding for enhanced nutrition and bio-active compounds in food legumes, Springer, Cham, pp 47–82
Parihar AK, Yadav R, Lamichaney A et al (2022a) Field pea breeding. In: Yadav DK, Dikshit HK, Mishra GP, Tripathi S (eds) Fundamentals of field crop breeding. Springer, Singapore, pp 1237–1321
Parihar AK, Kumar J, Gupta DS (2022b) Genomics enabled breeding strategies for major biotic stresses in pea (*Pisum sativum* L.). Front Plant Sci 13:861191
Parihar AK, Hazra KK, Lamichaney A et al (2022c) Characterizing plant trait (s) for improved heat tolerance in field pea (*Pisum sativum* L.) under subtropical climate. Int J Biometeorol 66(6): 1267–1281
Parihar AK, Hazra KK, Lamichaney A, Singh AK, Dixit GP (2023) Delineating the role of plant stature towards heat stress tolerance in field pea (*Pisum sativum* L.). Heliyon 9:e14539. https://doi.org/10.1016/j.heliyon.2023.e14539
Patto MV, Amarowicz R, Aryee ANA et al (2015) Achievements and challenges in improving the nutritional quality of food legumes. Crit Rev Plant Sci 34(1–3):105–143
Ramalingam A, Kudapa H, Pazhamala LT et al (2015) Proteomics and metabolomics: two emerging areas for legume improvement. Front Plant Sci 6:1116
Ranjbar Sistani N, Kaul HP, Desalegn G et al (2017) *Rhizobium* impacts on seed productivity, quality, and protection of *Pisum sativum* upon disease stress caused by *Didymella pinodes*: phenotypic, proteomic, and metabolomic traits. Front Plant Sci 8:1961
Rayner T, Moreau C, Isaac PG et al (2018) Genetic diversity and strategies for seed quality enhancement in *Pisum* (pea). In: Carlton R, Smith B, Topp K et al (eds) Aspects of applied biology 138: advances in legume science and practice. The Association of Applied Biologists, Wellesbourne, pp 141–148
Rehman AU, Shunmugam A, Arganosa G, Bett KE, Warkentin TD (2012) Inheritance of the lowphytate trait in pea. Crop Sci 52:1171–1175
Satoh H, Shibahara K, Tokunaga T et al (2008) Mutation of the plastidial α-glucan phosphorylase gene in rice affects the synthesis and structure of starch in the endosperm. Plant Cell 20:1833–1849
Shen S, Hou H, Ding C et al (2016) Protein content correlates with starch morphology, composition and physicochemical properties in field peas. Can J Plant Sci 96:404–412. https://doi.org/10.1139/cjps-2015-0231
Shivay YS, Singh U, Prasad R, Kaur R (2016) Agronomic interventions for micronutrient biofortifcation of pulses. Indian J Agron 61(4th IAC Special Issue):S161–S17
Sindhu A, Ramsay L, Sanderson LA et al (2014) Gene-based SNP discovery and genetic mapping in pea. Theor Appl Genet 127(10):2225–2241
Singh B, Singh JP, Shevkani K et al (2017) Bioactive constituents in pulses and their health benefits. J Food Sci Technol 54:858. https://doi.org/10.1007/s13197-016-2391-9
Sudheesh S, Lombardi M, Leonforte A et al (2015) Consensus genetic map construction for feld pea (*Pisum sativum* L.), trait dissection of biotic and abiotic stress tolerance and development of a diagnostic marker for the er1 powdery mildew resistance gene. Plant Mol Biol Rep 33: 1391–1403
Tahir R, Ellis PR, Bogracheva TY et al (2011) Study of the structure and properties of native and hydrothermally processed wild-type, lam and r variant pea starches that affect amylolysis of these starches. Biomacromolecules 12(1):123–133

Taran B, Warkentin T, Somers D et al (2003) Identification of quantitative trait loci for plant height, lodging resistance and reaction to mycosphaerella blight in pea (*Pisum sativum* L.) using an AFLP-based linkage map. Theor Appl Genet 107:1482–1491

Tar'an B, Warkentin T, Somers DJ et al (2004) Identification of quantitative trait loci for grain yield, seed protein concentration and maturity in field pea (*Pisum sativum* L.). Euphytica 136: 297–306. https://doi.org/10.1023/B:EUPH.0000032721.03075.a0

Tayeh N, Aluome C, Falque M et al (2015a) Development of two major resources for pea genomics: the geno Pea13.2K SNP array and a high density, high resolution consensus genetic map. Plant J. https://doi.org/10.1111/tpj.13070

Tayeh N, Klein A, Le Paslier MC et al (2015b) Genomic prediction in pea: effect of marker density and training population size and composition on prediction accuracy. Front Plant Sci 6:941

Thavarajah D, Ruszkowski J, Vandenberg A (2008) High potential for selenium bio-fortification of lentils (*Lens culinaris* L.). J Agric Food Chem 56:10747–10753

Tosh SM, Yada S (2010) Dietary fibres in pulse seeds and fractions: characterization, functional attributes, and applications. Food Res Int 43(2):450–460

Tzitzikas EN, Vincken JP, de Groot J et al (2006) Genetic variation in pea seed globulin composition. J Agric Food Chem 54(2):425–433. https://doi.org/10.1021/jf0519008

Urbano G, López-Jurado M, Aranda C et al (2006) Evaluation of zinc and magnesium bioavailability from pea (*Pisum sativum*, L.) sprouts. Effect of illumination and different germination periods. Int J Food Sci Technol 41(6):618–626

USDA (2020) U.S. national plant germplasm system. https://npgsweb.ars-grin.gov/gringlobal/descriptors

Van Berkel J, Conrads-Strauch J, Steup M (1991) Glucan-phosphorylase forms in cotyledons of *Pisum sativum* L.: localization, developmental change, in-vitro translation, and processing. Planta 185:432–439

Vigeolas H, Chinoy C, Zuther E et al (2008) Combined metabolomic and genetic approaches reveal a link between the polyamine pathway and albumin 2 in developing pea seeds. Plant Physiol 146:74–82. https://doi.org/10.1104/pp.107.111369

Wang TL, Hedley CL (1991) Seed development in peas: knowing your three 'r's'(or four, or five). Seed Sci Res 1(1):3–14

Wang TL, Smith CM, Cook SK et al (1987) An analysis of seed development in Pisum sativum. III. The relationship between the r locus, the water content and the osmotic potential of seed tissues in vivo and in vitro. Ann Bot 59:73–80

Wang K, Hasjim J, Wu AC et al (2014) Variation in amylose fine structure of starches from different botanical sources. J Agric Food Chem 62(19):4443–4453

Warkentin TD, Delgerjav O, Arganosa G, Rehman AU, Bett KE, Anbessa Y, Rossnagel B, Raboy V (2012) Development and characterization of low-phytate pea. Crop Sci 52:74–78

Weber H, Heim U, Golombek SD et al (1998) Expression of a yeast-derived invertase in developing cotyledons of Vicia narbonensisalters the carbohydrate state and affects storage functions. Plant J 16:163–172

Weckwerth W (2003) Metabolomics in systems biology. Annu Rev Plant Biol 54(1):669–689

Welham T, Domoney C (2000) Temporal and spatial activity of a promoter from a pea enzyme inhibitor gene and its exploitation for seed quality improvement. Plant Sci 159:289–299

White PJ (2016) Selenium accumulation by plants. Ann Bot 117(2):217–235

Winter P, Rubiales D, Fondevilla S (2016) Use of MACE technology to identify positional and expressional candidate genes for resistance to Didymella pinodes in pea. Second International Legume Society Conference

Zhang H, Jha AB, Warkentin TD, Vandenberg A, Purves RW (2018) Folate stability and method optimization for folate extraction from seeds of pulse crops using LC-SRM MS. J Food Compos Anal 71:44–55

Zhu ZP, Hylton CM, Rössner U et al (1998) Characterization of starch-debranching enzymes in pea embryos. Plant Physiol 118(2):581–590

Breeding Cowpea: A Nutraceutical Option for Future Global Food and Nutritional Security

Avi Raizada, Dhanasekar Punniyamoorthy, Souframanien Jegadeesan, Tesfaye Walle Mekonnen, and Penna Suprasanna

Contents

Abstract

Nutritional security has become the prime concern for world agriculture through improving the nutritional quality of crop plants and fostering nutri-rich crops. Malnutrition in the developing countries and increased occurrence of several

A. Raizada · D. Punniyamoorthy · S. Jegadeesan (✉)
Nuclear Agriculture & Biotechnology Division, Bhabha Atomic Research Centre, Trombay, Mumbai, India
e-mail: souf@barc.gov.in

T. W. Mekonnen
Department of Plant Sciences, University of the Free State, Bloemfontein, South Africa

P. Suprasanna (✉)
Amity Institute of Biotechnology, Amity University of Maharashtra (AUM), Mumbai, Maharashtra, India

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_26

health problems are now the major global challenges. Cowpea is a prime pulse legume grown predominantly in the tropical and subtropical regions of Asia, Africa, Latin America, and southern Europe. Compared to other legumes, cowpea is more resilient to climate change and exhibits wider adaptability in different agro-ecologies. Cowpea is a rich source of protein, carbohydrates, minerals, and vitamins with a low lipid content. Besides being nutritious, its health-promoting and health-protective effects are based on resistant starch, dietary fiber, phenolics, and peptides. Major health beneficial features of cowpea include anti-cancer, anti-diabetic, anti-inflammatory activities, and controlling blood lipid content. Germplasm evaluation is paving way for the identification of lines with high protein content and minerals (Cu, Fe, Zn, Mg, Ca, and K) which could be used in breeding for new biofortified cowpea cultivars. Development of databases such as "EDITS-Cowpea" for enabling exploration of cowpea traits, especially those related to grain quality-related traits and "Cowpea Genomics Initiative" for applying modern molecular genetic tools for gene discovery will foster research aimed at cowpea improvement. This chapter examines the nutritive value of cowpea with more emphasis on the remarkable nutraceutical properties of cowpea and suggests taking cowpea forward as a future smart crop for tackling global food and nutritional security.

Keywords

Pulses · Cowpea · Nutritional security · Nutraceuticals · Human health

1 Introduction

Food security for the increasing world population has become a main concern for agricultural scientists and plant breeders, and it is estimated that food production will have to be enhanced by 70% by 2050 (Fróna et al. 2019). The problem is also compounded by the nutritional deficiencies and malnutrition among the world population (FAO 2017). This necessitates that the food will have to be produced in abundance and of good nutritional quality to ensure health of the future generations. Plant breeding over the past several decades has made a significant contribution to the development of high yielding and good quality varieties of cereals, pulses, and oil seeds besides other economically important plants to meet food and nutritional security (Mir et al. 2021).

Cowpea (*Vigna unguiculata*), also known as China bean, black-eyed bean, black-eyed pea, and southern pea, is an annual bean plant and member of the family Fabaceaeor Leguminosae (Oyewale and Bamaiyi 2013). It is widely cultivated in the tropical regions of the world such as Southeast Asia, Africa, Southern United States, and Latin America due to its ability to tolerate climate change. Cowpea is majorly produced and consumed in Sudano–Sahelian zone of sub-Saharan Africa (Boukar

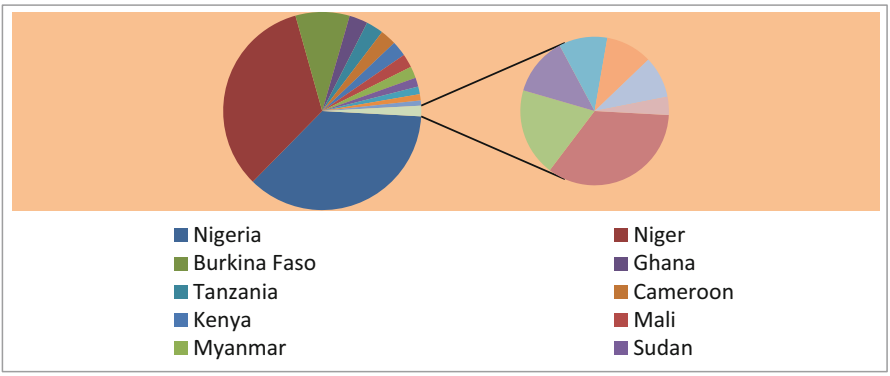


Fig. 1 Global cowpea production (FAOStat 2020)

et al. 2019). According to FAOSTAT report, worldwide production estimates over 8.9 million metric tons of dry cowpeas in 2020, of which 86% is contributed by Western Africa, mainly Nigeria and Niger with 6.3 million tons (FAOSTAT 2020) (Fig. 1). At the global level, cowpea production and yield have been increased by 88% and 35%, respectively (Nedumaran et al. 2015). Most of the cowpea plant parts are edible and consumed such as juvenile leaves, growing points, raw/immature pods, green seeds, and dried/desiccated seeds (Gerrano et al. 2019). It is grown for its high nutritious value, and recently there has been focus on the nutraceutical properties and as fodder for livestock. Moreover, due to its nutritious leaves, cowpea is placed in class of the chief vegetables in Africa and Asia (Mohammed et al. 2021). The American Pulse Association declared "pulses" as the most versatile and multi-faceted food source in the world (American Pulse Association 2020). It is often referred to as a multi-potential crop for the future (Fig. 2).

Cowpea is reported to have originated from Africa (Lazaridi et al. 2016). A recent study in 2020 (Herniter et al. 2020) based on genetic, textual, and archeobotanical data proposed a likely spread of cowpea from the two centres of domestication, West Africa and East Africa. From West Africa cowpea was spread by the Bantu migrations south to the equatorial rainforest and then to the areas of modern Sudan, South Sudan, and Ethiopia followed by diversion into three branches. First branch leads to Southern Africa, which joined the East African domestication. Second branch directed toward north up to the Nile and Egypt and remained there up to 2500 BCE. Later on, by 400 BCE, this branch had entrenched itself as a key food crop in the Mediterranean basin and moved to Spain's colonial holdings in the New World, including the modern south-western United States. The third branch makes its way to the west coast of India by 1500 BCE through the "Sabaean lane" in modern Yemen and then spreads to Southeast Asia. *Vigna unguiculata* ssp. *unguiculata* var. *spontanea* is assumed to be the wild ancestor of cultivated cowpea that is grown in sub-Saharan Africa (Pasquet et al. 2021).

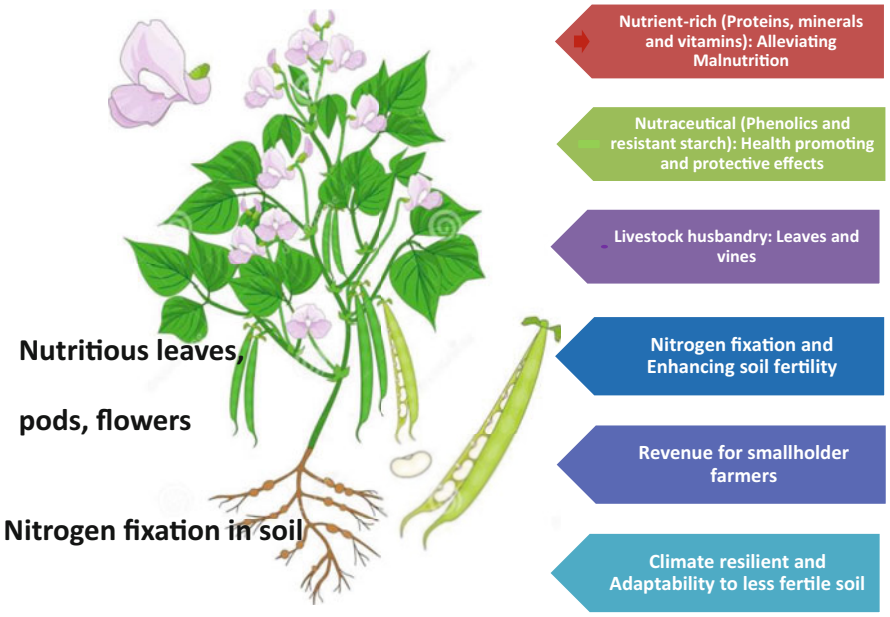


Fig. 2 Multifaceted potentials of cowpea

2 Genetic Resources and Genetic Diversity

Crop genetic resources including germplasm collections are essential for national and global agricultural security. Genetic diversity of crops is important for sustainable development and food security because this gene reservoir will aid in the future in further improvements in the elite cultivars with regard to better performance and well-adaptedness (Pathirana and Carimi 2022). To date several genetic resources have been developed in cowpea to aid in breeding elite varieties. The different forms of these developed genetic resources include physical and genetic linkage maps, genome sequences, databases, microarrays, molecular markers, etc. Few examples of different genetic resources in cowpea are genotyping assays and genetic maps based on single nucleotide polymorphism (SNP), physical maps, mapped quantitative trait loci (QTLs) traits, consensus genetic maps of cowpea (González et al. 2016), reference genome sequence of cowpea (Lonardi et al. 2019), cDNA sequences, unigenes, genic-SSR markers (Mahalakshmi et al. 2007), and linkage maps for cowpea developed using molecular markers and their further refinements through advanced markers (Muchero et al. 2009). Several genetic diversity studies have been conducted to investigate the evolutionary relationships among different genotypes, relationships with wild accessions, origin, taxonomy, domestication, and evolutionary pattern in cowpea. Initially studies were performed using conventional parameters such as allozymes and seed storage proteins that was followed by

different marker systems such as chloroplast DNA polymorphism, restriction fragment length polymorphism (RFLP), amplified fragment length polymorphisms (AFLP), DNA amplification fingerprinting (DAF) simple sequence repeats (SSRs), cross species SSRs from *Medicago*, inter-simple sequence repeats, sequence tagged microsatellite sites (STMS), and single-nucleotide polymorphism (SNP) markers.

The largest collection of cowpea consisting of 16,569 cultivated accessions from 100 countries and around 1500 wild accessions of *vigna* is maintained by the International Institute of Tropical Agricultural (IITA). Based on geographical, agronomical, and botanical descriptors, a central group of 2120 accessions and a minor collection of 376 accessions have been established at IITA (Boukar et al. 2019). Other major collections with overlapped data include the US Department of Agriculture–Agricultural Research Service (USDA-ARS) (Griffin, Georgia, USA) with 8379 accessions, the National Bureau of Plant Genetic Resources (NBPGR, New Delhi, India) holding 4003 accessions, and the University of California, Riverside (UCR, California, USA) harboring 5000 accessions and a mini core of 368 accessions. These cowpea accessions exhibited variations among each other with context to several morphological and agronomical traits such as plant pigmentation, plant kind, plant height, leaf type, growth habit, photosensitivity or insensitivity, maturity, nitrogen fixation, fodder grade, tolerance to high temperature and water deficit, root architecture, pod features, seed traits, grain quality and reaction/response to diseases, root-knot nematodes, insect pests (aphids, bruchid, thrips), and parasitic weeds (Boukar et al. 2020). Large gene bank collections include those of IITA (Nigeria), USDA (Southern Regional Plant Introduction Station, Georgia), World Vegetable Centre (Taiwan), and the N.I. Vavilov Research Institute of Plant Industry (Russia). Padhi et al. (2022) explored 120 diverse cowpea germplasm lines to search for nutri-dense genotypes using biochemical traits and observed broad variability for protein content (19.4 to 27.9%), starch (27.5 to 42.7 g 100 g^{-1}), amylose (9.65 to 21.7 g 100 g^{-1}), TDF (13.7 to 21.1 g 100 g^{-1}), and TSS (1.30 to 8.73 g 100 g^{-1}). The study suggested that the collection showed some nutrient-dense lines having more than a single trait with high nutritional potential.

3 Nutritional and Nutraceutical Profile

The world is facing major critical situations of malnutrition in downtrodden population of developing countries and prevalence of chronic diseases in well-off people in developed countries. Protein energy malnutrition (PEM) is a very serious public health issue in many less developed nations (Bessada et al. 2019). Globally, one-third of all child deaths were estimated to be due to malnutrition, of which 54% occurred in underdeveloped countries (Bain et al. 2013). Thus, there is an urgent need to identify foods that are both nutritious and have nutraceutical properties and to introduce these food types into our regular diet so that their health-protective and health-promoting effects will help in controlling frequent cases of several chronic diseases.

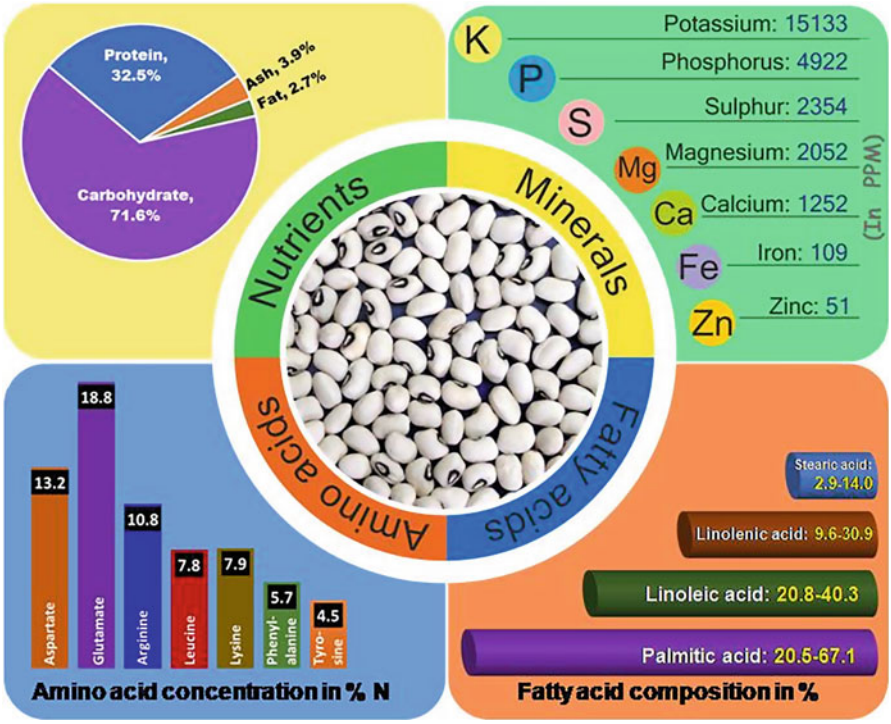


Fig. 3 Major nutritional composition of cowpea seeds (based on the data from Gonçalves et al. 2016)

Legumes are well-known for their protein-richness, of which cowpea has gained more attention due to its remarkable nutritional profile and nutraceutical properties that make it unique among other pulses (Dhanasekar et al. 2021) (Fig. 3). Cowpea is known as "the poor man's meat" because of its protein-rich nature complemented by its less expensive and affordable access to rural poor people (Dugje et al. 2022), and also because the protein content is approximately equal to certain meat types (18–25%). It also has digestible and non-digestible carbohydrates, potassium, and very low lipid and sodium content. The composition of different nutritional components in grains is given in Fig. 4. Cowpea leaves are rich in micronutrients, nutraceuticals, antioxidants (alpha tocopherols, flavonoids, lycopene), and anti-proliferating compounds (Owade et al. 2020). Dakora and Belane (2019) suggested that cowpea leaves could meet the suggested day-to-day dietary intake requirement of the micronutrients Fe and Zn by consuming 4 mg and 76 mg of leaf on a dry matter basis. Cowpea leaves are a nutritious food source, with an abundance of protein and minerals, digestible and non-digestible carbohydrates, and potassium but low lipids and sodium (Kamara et al. 2010). Leaves show higher nutritional content than grains (Fig. 5).

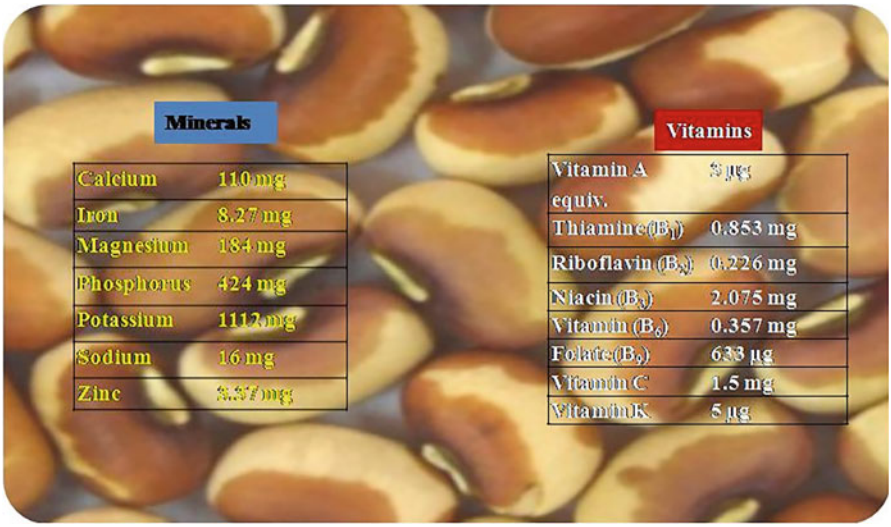


Fig. 4 General nutritional profile per 100 g of raw cowpea seeds. (Source: USDA nutrient database)

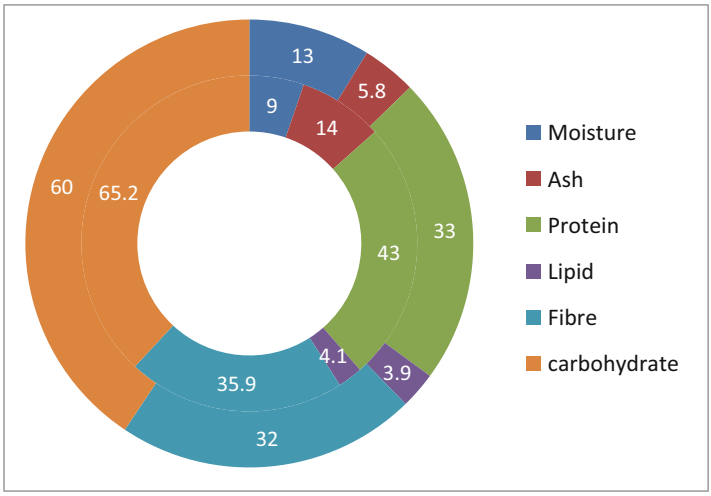


Fig. 5 Proximate and fiber composition (%) of leaves and grain (adopted from Mekonnen et al. 2022)

3.1 Protein Profile

Cowpea is consumed as a high-quality plant-based protein source (Jayathilake et al. 2018) with a protein content of 27 to 43% in leaves and 21 to 33% in dry grains (Gerrano et al. 2019). Cowpea protein fraction includes globulins (legumin and

Table 1 Amino acid composition of cowpea grain and leaves (adopted from Mekonnen et al. 2022)

Amino acid	Leaves (g/100 g protein)	Grain (g/100 g protein)	Amino acid	Leaves (g/100 g protein)	Grain (g/100 g protein)
Arginine	7.4–17.3	5.0–10.8	**Lysine**	3.0–16.3	3.5–8.0
Aspartic acid	10.8–26.7	6.0–13	**Methionine**	1.0–4.5	0.9–3.5
Alanine	4.2–9.8	3.4–5.1	**Phenylalanine**	4.6–14.4	4.4–9.9
Cysteine	0.5–2.9	0.3–2.4	**Proline**	4.0–15.9	3.1–8.9
Glutamic acid	17.2–45.3	8.5–19	**Serine**	3.0–11.6	3.8–5.8
Glycine	3.8–12.6	3.1–4.8	**Threonine**	3.2–10.8	3.0–5.9
Histidine	1.8–8.6	2.0–4.41	**Tryptophan**	1.3–4.1	0.9–1.5
Isoleucine	4.1–11.1	2.8–5.4	**Tyrosine**	3.0–9.3	2.6–4.5
Leucine	7.4–19.6	5.7–11.3	**Valine**	5.0–12.8	3.4–6.2

vicilin/β-vignin), albumins, glutelins, and prolamins (Santos et al. 2012). Because of this, cowpea has been highly promoted in economically backward regions to control protein malnutrition (Iqbal et al. 2006).

Cowpea leaves and grains show a rich profile of different amino acids including essential amino acids like valine, leucine, phenylalanine, lysine, and tryptophan. The amino acid profile of leaves and grains is presented in Table 1.

3.2 Minerals and Vitamins

Cowpea seeds, leaves, and beans are enriched with vital minerals of both macronutrients (Ca, K, Mg, P and S) and micronutrients (Cu, Fe, Mn, Zn, Na, Al, Se, B) that are required for proper functioning of human body (Owade et al. 2020). Cowpea is a source of different vitamins, of which most prevalent are vitamin A, C, and B complex (thiamine, riboflavin, pantothenic acid, pyridoxine, and folic/folate acid) and gamma tocopherol. Cowpea leaves have more vitamin C and minerals compared to grains (Mekonnen et al. 2022). The major and micro-mineral profiles of leaves, immature pods, and grains are presented in Table 2.

3.3 Lipids and Fatty Acids

Cowpea is a low lipid grain crop compared to other legumes (chickpea, lentil, green gram, and lupin) (Belane and Dakora 2011) with 0.5% to 3.9% lipid in grain and 1.3 to 4.3% in leaves. The lipid profile of cowpea is composed of triglycerides (41.2%), phospholipids (25.1%), monoglycerides (10.6%), free fatty acids (7.9%), diglycerides (7.8%), sterols (5.5%), hydrocarbons, and sterol esters (2.6%)

Table 2 Mineral composition of cowpea grains, immature pods, and leaves (adopted from Mekonnen et al. 2022)

Minerals	Grains Mean range	Immature pods Mean range	Leaves Mean range
Macro-minerals (mg/100 g dry matter)			
Phosphorus	2.3–6.10	383.43–537.53	2.1–592.4
Potassium	9.30–35.60	170.74–240.78	9.57–1445.2
Magnesium	4.3–8.4	297.97–426.20	1.3–227.4
Sulfur	153.3–200.0		120.0–147.3
Micro-minerals (mg/100 g dry matter)			
Copper	0.15–2.2	0.48–0.95	0.5–2.2
Iron	26.76–182.33	6.01–9.78	3.4–10.6
Manganese	10.57–204	2.11–4.77	1.38–4.3
Sodium	11.59–43.95	13.70–32.93	8.4–79.81
Zinc	2.78–22.3	1.42–5.63	2.4–5.11
Aluminum		1.84–7.86	
Boron	3.14–5.01	2.13–4.03	1.47–2.14
Selenium		2.5–3.4	

(Kapravelou et al. 2015). Among fatty acids, palmitic and linoleic acids and, within sterols, stigmasterol (42.1 to 43.3%) are predominant (Antova et al. 2014).

3.4 Carbohydrates

Cowpea is also rich in carbohydrates containing 30.39 to 31.11% in leaves and 50 to 60% in grains. Popova and Mihaylova (2019) suggested that the good fraction of carbohydrates consists of sucrose, glucose, fructose, galactose, and maltose, whereas anti-nutrient components of carbohydrates are mostly raffinose, stachyose, and verbascose.

4 Health-Promoting and Health-Protective Properties

Besides above-specified nutritional value, cowpea exerts several health benefits due to presence of soluble and insoluble dietary fibers, phenol-derived compounds, other functional agents, anthocyanins, and carotenoids (de Silva et al. 2021). Epidemiological evidences showed the nutraceutical aspects, i.e., health-promoting and disease-preventing effects of cowpea such as protection against many incurable and immedicable health situations such as cardiovascular diseases, hypercholesterolemia, and obesity (Frota et al. 2015), anti-diabetic (Barnes et al. 2015), anti-cancer (de Silva et al. 2021), anti-inflammatory (Awika and Duodu 2017), antihypertensive and hypocholesterolemia (Tadele 2019), reducing plasma low-density lipoprotein (Talabi et al. 2022), gastrointestinal disorders (Khalid II and Elharadallou 2012),

weight loss (Perera et al. 2016), and improving assimilation and strengthening blood flow (Trehan et al. 2015). The nutraceutical aspects of cowpea were reviewed (Jayathilake et al. 2018) and credited with plant-derived compounds, resistant starch, dietary fiber, less fat, and good unsaturated fatty acids. Interestingly, upon germination, the nutritive profile of cowpea seeds is enhanced as observed by increase in antioxidant capacity, vitamin C (Doblado et al. 2007), β-carotene, phenolics (hydroxycinnamic acid, syringic acid, vanillaldehyde, ferulic acid, sinapic acid, *p*-coumaric acid, benzoic acid, ellagic acid, and cinnamic acid), and flavonoid content.

4.1 Protein Hydrolysates and Peptides

The protein lysates of cowpea vegetate shows a low lysine/arginine proportion like soybean, thus making it a potential functional ingredient for reducing cholesterol (Kanetro 2015). Cowpea bioactive compounds such as peptides that are products of enzymatic hydrolysis or fermentation are reported to create favorable physiological conditions for proper functioning of the human body (Marques et al. 2015). These peptides function by acting as antihypertensive (Boonla et al. 2015), anti-dyslipidemic (Udenigwe and Rouvinen-Watt 2015), antioxidative (Marques et al. 2015), anti-carcinogenic and antimicrobial (Felicio et al. 2017), and anti-diabetic (Barnes et al. 2015). Cowpea peptides are known to prevent the occurrence of diabetes mellitus by imitating the activity of insulin and inhibiting dipeptidyl peptidase IV activity (Barnes et al. 2015). The antioxidative properties of cowpea peptides were attributed to the hydrophobic and aromatic amino acids like leucine, isoleucine, tyrosine, phenylalanine, tryptophan, and the sulfur-bearing amino acid cysteine due to their proton giving property to free radicals (Xiong et al. 2013). Similarly cowpea proteins and peptides showed antihypertensive by inhibiting angiotensin-converting enzyme (ACE) (de Leon et al. 2013) and hypo-cholesterolemic effects occur in several ways such as bile acid-binding, disruption of cholesterol micelles, changing hepatic and adipocytic enzyme actions, and gene expression of lipogenic proteins, as well as by inhibiting HMG-CoA reductase activity (Marques et al. 2018).

4.2 Phenolics

Cowpea predominantly contains phenolics (70% free phenolics and 30% bound phenolics), flavonoids (flavonols and flavan-3-ols), coumaric acid and ferulic acid (seed), gallic acid, protocatechuic acid, and *p*-hydroxybenzoic acid (seed coat) and thus is supposed to exert high antioxidant activity (Gutierrez-Uribe et al. 2011). Anthocyanins found in cowpea are delphinidin-3-*O*-glucoside, cyanidin-3-*O*-glucoside, delphinidin-3-*O*-galactoside, cyanidin-3-*O*-galactoside, petunidin-3-*O*-glucoside, peonidin-3-*O*-glucoside, and malvidin-3-*O*-glucoside (Ha et al. 2010). Many studies have reported the phenolic composition and their functioning mechanisms in different cowpea varieties (Liyanage et al. 2014). Phenolic compounds have

hypocholesterolemic activity because of inhibition of oxidation of lipids, lowering of blood triglycerides, total cholesterol, LDL, and surge of blood HDL (Hachibamba et al. 2013), anti-inflammatory by downregulating pro-inflammatory gene expression (Ojwang et al. 2015), and having an anticancer effect as antioxidants shields DNA from oxidation and suppress cancerous cell division (Hachibamba et al. 2013).

4.3 Resistant Starch and Fiber

Coming to the other health benefits, cowpea contains high amount of resistant starch up to 12.65 gm per 100 gm (Eshwarage et al. 2017; Chen et al. 2010) which makes it a low glycemic index food. As mentioned earlier, resistant starch and dietary fiber have antidiabetic effect due to slow release of glucose (Onyeka 2007) and hypocholesterolemic effect due to diminution of bile acids from the circulation, abatement of converting cholesterol to supplemental bile acids, and augmenting expulsion of fecal fat (Perera et al. 2016). The two major health benefits of resistant starch are that firstly, it slows the rate of digestion, thus slowing release of glucose into the body followed by less uptake of glucose by the intestinal cells and secondly, due to their incomplete digestion by human digestive enzymes, they act as a substrate for colonic microbes (including probiotics) resulting in production of short-chain fatty acids (butyrate) that aid in proper lipid function and cancer prevention. Cowpea also has low calorific value and thus helps with glucose regulation in diabetic patients and better weight control for the obese (Oboh and Agu 2010).

Cowpea has both soluble and insoluble high fiber content and hence has associated health advantages (Eshwarage et al. 2017). Soluble fiber helps in regulating blood cholesterol and glucose levels, while insoluble fiber due to its water/moisture retention property helps in smooth passage of waste materials through intestine and colon, thereby preventing haemorrhoids, constipation, many other digestive difficulties, colon cancer, diabetes, obesity, cardiovascular diseases, and numerous other long-term health complications (Eshwarage et al. 2017). Some additional health benefits of cowpea include eliminating urination problems such as uneasiness or obstructions, managing leucorrhea, or abnormal vaginal discharge (Alfa et al. 2020).

4.4 Anti-nutritional Factors

Although numerous studies claim cowpea is highly nutritious and a good nutraceutical food, its consumption is still limited due to presence of several anti-nutritional factors, poor digestibility, and lack of sulfur-containing amino acids. These anti-nutrients include some phenolic compounds, such as proanthocyanidins (Ojwang et al. 2013), phytic acid, tanins (Lattanzio et al. 2005), hemagglutinins (Aguilera et al. 2013), cyanogenic glucosides, oxalic acid, dihydroxyphenylalanine and saponins, and enzyme inhibitors (protease inhibitors, phytocystatins) (Monteiro et al. 2017). Phenolic compounds bind proteins and chelate divalent metal ions (Ojwang et al. 2013). Phytic acid (PA), an anti-nutritional factor, is known to conjugate phosphorus and

other essential elements like iron and make it unavailable to organisms that feed on seeds rich in PA. Mutations affecting PA content have been identified, and low PA mutants have been isolated in pulses including cowpea (Dhanasekar and Reddy 2017). These mutants are being exploited in varietal development program so as to reduce the PA content to reasonable levels without affecting the physiological balance as they are further involved in responses to biotic and abiotic stresses (Dhole and Reddy 2016; Dhanasekar and Reddy 2017). Another important anti-nutritional factor in cowpea is the raffinose family oligosaccharides (RFOs) known to cause flatulence in organisms ingesting cowpea seeds. The RFOs are highly recalcitrant to various processing methods, and genetic means of reducing the content is the only amenable method. Therefore, it becomes imperative to identify genotypes with low RFOs content. Mutants with low RFOs have been identified in cowpea (Dhanasekar and Reddy 2015) that could be potential donors for developing varieties with low RFOs. Nevertheless, appropriate processing techniques can be used to lower many of the anti-nutritional compounds and enhance their bioavailability levels. Knowledge of nutritional, nutraceutical, and anti-nutritional aspects of cowpea will help in designing appropriate dietary plans/guidelines as per ethnic groups and geographic regions. Besides being in high demand due to its nutritious and nutraceutical values, cowpea cultivation is also supported because of the desirable agronomic attributes such as ease of cultivation, less necessity for fertile soils, their adaptability, and steadiness across all continents, even in drought-afflicted regions (de Silva et al. 2021).

5 Conventional and Molecular Approaches for Enhancing Nutritional Potential

Different processing methods, depending on geographical regions, are widely adopted by respective natives to reduce anti-nutrients present in cowpea and to enhance nutritional profile and also for the ease of intake. These methods include boiling, sprouting, steaming, frying, soaking, de-hulling, and grinding, which result in the alteration of the properties and bioavailability of some nutrients, increase in protein and mineral content (Fabbri and Crosby 2016), increase in phenolics and flavonoids (Laila and Murtaza 2014), and protein quality and digestibility (Deol and Bains 2010). Fermented cowpea flour showed improved antioxidant and hypolipidemic effects on rat (Kapravelou et al. 2015).

Extensive efforts through molecular tools have been made for breeding cowpea varieties resistant to biotic and abiotic stresses, to enhance yield and productivity, but reports on improving nutritional profile through molecular approaches are scarce. However, research on enhancing cowpea nutritional aspects has been conducted for evaluating biochemical properties of cowpea germplasm. Considerable data has been generated on estimation of the protein and minerals content of cowpea germplasm with the objective to identify appropriate parents for breeding nutrient-dense improved varieties (Fig. 3). For example, evaluation of cowpea germplasm lines for protein and minerals (Cu, Fe, Zn, Mg, Ca, and K) content (Gerrano et al. 2019) and analysis of cowpea cultivars for proteins and minerals under rain-fed conditions in

Petrolina, Brazil (Santos and Boiteux 2013), suggested the identification of high protein and mineral content genotypes which could be used in breeding for new biofortified cowpea cultivars.

Reports on cowpea genetics for biofortification with crucial minerals are still lacking, of which one study provided genetic factors for enhancing minerals in cowpea seeds. This study claimed the least number of genes controlling the augmentation of minerals ranging from 2 (K) to 11 (P) with transgressive segregation pattern and either oligogenic or polygenic control for all minerals analyzed (Fernandes et al. 2015). Composite interval mapping detected two QTLs for total soluble solid in pods using the population derived from the cross between yardlong bean (accession JP81610) and a wild cowpea (*V. unguiculata* ssp. *unguiculata* var. *spontanea*) (accession JP89083) (Kongjaimun et al. 2013). Recently in a genotype-by-environment interaction study, the expression of nutritional properties (protein and minerals concentrations) in cowpea leaves was assessed in different agro-ecologies of South Africa and typical agronomical practices of smallholder farmers (Gerrano et al. 2022). This study showed genetic variations among selected genotypes for all four analyzed traits and also influence of climate on expression of these traits. This study suggested that nutritional profile of legume plants is a function of local soil properties and soil health (Gerrano et al. 2022).

Mutation breeding is generally adopted to introduce desired simply inherited trait in elite cultivars. The International Atomic Energy Agency (IAEA) in association with the Food and Agriculture Organization (FAO) encourages the deployment of mutation-inducing technologies in plants for its member states. So far, there have been 22 mutant varieties developed using physical and chemical mutagens in cowpea. Few reports are available in cowpea cultivar development for biotic and abiotic stress tolerance using induced mutations (Horn et al. 2017), of which one study showed increase in the protein content in grains of some mutants by up to 13.3% (mutant from IT84E-124) and 13.64% (mutant from Vita 7) upon treatment with 1.0 mM NaN_3 (Odeigah et al. 1998). Raina et al. (2022) reported that there was a concurrent increase in yield and nutrient density (Protein, Fe, Zn, and Cu) in M_4 mutant lines in cowpea. Seven cowpea cultivars developed through mutation breeding exhibiting high seed productivity, earliness, large grain size, resistance to yellow mosaic virus, or augmented fodder production were released between 1981 and 2007 in India (Punniyamoorthy et al. 2007). One of the mutant varieties, the multifaceted cowpea mutant variety "TC-901," has desirable attributes like high grain yield, fodder yield, high seed protein content (28%), resistance to cowpea mosaic virus, and amenable for summer cultivation. The progress in enhancing nutritional profile of cowpea is underway through evaluation of germplasm lines for identifying appropriate parents for breeding elite, nutri-rich cultivars.

6 Genomics of Nutritional Quality

Information on genomics of the crop nutrition profile has laid a starting point for the implementation of advanced molecular breeding and genetic engineering methods for crop improvement and also helps in switching from time-consuming

labor-intensive conventional breeding. Nutritive value of a crop can be enhanced either by upregulating genes/QTLs associated with nutritive traits or by suppressing genes involved in anti-nutrient biosynthesis for which knowledge of target genes is required. Efforts have been made to explore and understand the genomics behind nutritional features in crops like cereals, pulses, oil seeds, legumes, millets, and vegetables in the form of evaluation of crop germplasm, development of genetic resources through introduction of new genotypes/novel genes, mapping and characterization of genes through quantitative trait locus (QTL) interval mapping and sequencing, association mapping such as genome-wide association study (GWAS), and marker-assisted breeding as discussed as follows. For example, development of sorghum with low cyanogenic potential suitable for cattle feed generated by down-regulating an important enzyme of dhurrin biosynthesis pathway through antisense strategy (Pandey et al. 2019) provides opportunities for fine-tuning nutritional quality in grain crops. Interestingly in mung bean which is close to cowpea, 43 noteworthy marker trait associations (MTAs) for seed calcium, iron, potassium, manganese, phosphorous, sulfur, or zinc concentrations were discovered through genotyping by sequencing (GBS) approach (Wu et al. 2020). Coming to the crop improvement through genome engineering, the technique has been successfully applied to few crops including soybean for reducing linolenic acid (silencing of the ω-3 *fad3* gene) (Flores et al. 2008) and increasing oleic acid (suppressing *fad2-1* gene) (Christou et al. 1990). A database for cowpea, "EDITS-Cowpea," has all the required information on the cowpea traits especially related to grain quality-related traits which can be useful to breeders for crop improvement (EDITS-Cowpea 2022). A search conducted for nutritional quality as trait and zinc content as the specific search item showed that the cowpea varieties (240) in the database have a range of zinc content (34.5–46 mg/g).

A remarkable initiative is taken by the Kirkhouse Trust, a UK-based charitable organization through Cowpea Genomics Initiative (CGI) project (http://cowpeagenomics.med.virginia.edu/CGKB/) to help cowpea research community. CGI project is aimed at leveraging advanced molecular tools for gene study and bettering cowpea. This project attempts omics studies including transcriptome, proteome, and metabolome analyses to get comprehensive knowledge on the fundamental biology of host and important agronomic characteristics and also sequencing and annotation of the gene space (gene-rich region of the cowpea genome) (Chen et al. 2007). Muñoz-Amatriaín et al. (2017) developed genome resources for the analysis of an African cultivar IT97K-499-35 which included whole-genome shotgun assembly, a bacterial artificial chromosome physical map, and assembled sequences for use in linkage mapping, synteny analysis, and germplasm characterization. In a further study, Lonardi et al. (2019) developed a genomic assembly with the help of single-molecule real-time sequencing complemented with optical and genetic mapping tolls to categorize repetitive elements, genes, and gene families. Noteworthy advanced genome editing technology, CRISPR-Cas9 system, has been successfully applied first time in cowpea which involved the inactivation of symbiosis receptor-like kinase gene via *Agrobacterium*-mediated hairy root transformation method (Ji et al. 2019). Such

project initiatives and implementation of genomics techniques may help in promoting cowpea research for improvement in nutritional quality in related crops through genomics-based trait introgression.

7 Future Perspectives and Conclusions

Increased food production and management of malnutrition always remain major challenges for the underdeveloped and developing countries, whereas in the developed world, increase in the occurrence of several chronic diseases and occupation-related hazards are a priority. The situation is getting worrisome due to the continuously increasing population and climate changes that directly affect agriculture sector. Cultivation of important staple crops in all ecological conditions and climate change scenarios is not possible. In this regard, crops with wider adaptability can help in ensuring food and nutritional safety. Cowpea is a drought-tolerant, climate-adaptable crop amenable to diverse cropping systems. Cowpea is an excellent nutrient-rich food source belonging to orphan crop category that remains to be extensively cultivated since, besides being highly nutritious, it also possesses many health-promoting and health-protective effects. Cowpea nutraceutical properties are attributed to bioactive or functional chemicals like peptides, resistant starch, digestible fiber, plant-derived compounds, antioxidants, vitamins, etc. that get better depending on processing methods. Cowpea has been shown to better the lipid profile, blood glucose content, blood pressure, cancer prevention, anti-inflammatory, anti-diabetic, etc. It is a multipurpose legume crop used for both human consumption and livestock fodder. In the current scenario, though cowpea is a nutrient dense food with several health benefits, it is a neglected crop because of huge losses in its production due to biotic and abiotic stresses, cultural beliefs, and limited research priority. Modern breeding technological interventions will have to be accelerated for cowpea breeding. Also, intensive clinical research on the anti-inflammatory and anti-cancer activity of cowpea is required to realize the nutraceutical aspects of cowpea for better acceptability.

Breeding for biofortification in cowpea is in the budding stage, and hence the adoption of molecular tools and advanced genomic strategies is needed to accelerate the progress of development of nutrient-dense and nutraceutical-rich varieties. This could be achieved through mining available genetic resources for target and novel genes/traits in association with chance breeding (targeted mutation, hybridization, backcrossing, pedigree, and recurrent selection), modern breeding methods (space breeding, speed breeding, genomic selection, and gene/genome editing), innovative techniques (mutagenesis breeding), transgenic development, demand-led breeding, and multi-omics analysis studies. Cowpeas can become a super crop for alleviating nutrient deficiency and health problems. With further research on its nutraceutical aspects, its acceptance will also be on the rise. Being the cheapest protein source and climate change-resilient crop, next-generation cowpea could be promoted for achieving future food and nutritional security at the global level.

References

Aguilera Y, Diaz MF, Jimenez T, Benitez V, Herrera T, Cuadrado C et al (2013) Changes in nonnutritional factors and antioxidant activity during germination of nonconventional legumes. J Agric Food Chem 61:8120–8125

Alfa AA, Tijani KB, Omotoso OD et al (2020) Nutritional Values and Medicinal Health Aspects of Brown, Brown-Blackand White Cowpea (*Vigna unguiculata* L. Walp.) Grown in Okene, Kogi State, Nigeria. Asian J Adv Res Reports 14:114–124

American Pulse Association (2020) Meet Pulses: The World's Most Versatile Superfood. Available online: https://www.usapulses.org/consumers/resources/delicious

Antova GA, Stoilova TD, Ivanova MM (2014) Proximate and lipid composition of cowpea (*Vigna unguiculata* L.) cultivated in Bulgaria. J Food Compos Anal 33:146–152

Awika MJ, Duodu GK (2017) Bioactive polyphenols and peptides in cowpea (*Vigna unguiculata*) and their health promoting properties: a review. J Funct Foods 38:686–697

Bain LE, Awah PK, Awah KP, Geraldine N (2013) Malnutrition in sub-Saharan Africa: burden, causes and prospects. Pan Afr Med J 15:120

Barnes M, Uruakpa F, Udenigwe C (2015) Influence of cowpea (*Vigna unguiculata)* peptides on insulin resistance. J Nutr Health Food Sci 3:1–3

Belane AK, Dakora FD (2011) Levels of nutritionally-important trace elements and macronutrients in edible leaves and grain of 27 nodulated cowpea [*Vigna unguiculata* (L.)Walp.] genotypes grown in the Upper West Region of Ghana. Food Chem 125:99–105

Bessada SM, Barreira JC, Oliveira MBP (2019) Pulses and food security: dietary protein, digestibility, bioactive and functional properties. Trends Food Sci Technol 93:53–68

Boonla O, Kukongviriyapan U, Pakdeechote P, Kukongviriyapan V, Pannangpetch P, Thawornchinsombut S (2015) Peptides-derived from Thai rice bran improves endothelial function in 2K-1C renovascular hypertensive rats. Nutrients 7:5783–5799

Boukar O, Belko N, Chamarthi S, Togola A, Batieno J, Owusu, et al. (2019) Cowpea (*Vigna unguiculata*): genetics, genomics and breeding. Plant Breed 138(4):415–424

Boukar O, Abberton M, Oyatomi O, Togola A, Tripathi L, Fatokun C (2020) Introgression Breeding in Cowpea [*Vigna unguiculata* (L.) Walp.]. Front Plant Sci 11:567425

Chen X, Laudeman TW, Rushton PJ et al (2007) CGKB: an annotation knowledge base for cowpea (*Vigna unguiculata L.*) methylation filtered genomic gene space sequences. BMC Bioinform 8:129

Chen L, Liu MR, Qin C, Meng Y, Zhang J, Wang Y et al (2010) Sources and intake of resistant starch in the Chinese diet. Asia Pac J Clin Nutr 19:274–282

Christou P, McCabe DE, Martinell BJ, Swain WF (1990) Soybean genetic engineering –commercial production of transgenic plants. Trends Biotechnol 8:145–151

Dakora FD, Belane AK (2019) Evaluation of protein and micronutrient levels in edible cowpea (*Vigna Unguiculata* L. Walp.) leaves and seeds. Front Sustain Food Syst 3:70. https://doi.org/10.3389/fsufs.2019.00070

De Leon RC, Torio MAO, Manalo MN, Aguda RM (2013) Isolation, purification and characterization of the major storage protein in cowpea (*Vigna unguiculata*) seed with bioactive peptides exhibiting antiangiotensini-converting enzyme activity. 42nd Annual Convention of the Kimika Ng Pilipinas-Southern Tagalog, University of the Philippines Los Banos, Laguna, pp. 26

de Silva AC, de Freitas Barbosa M, de Silva PB, de Oliveira JP, de Silva TL, Junior DLT, de Moura Rocha M (2021) Health benefits and industrial applications of functional cowpea seed proteins. In: Jimenez-Lopez JC (ed) Grain and seed proteins functionality. IntechOpen. https://www.intechopen.com/chapters/75744,, London. https://doi.org/10.5772/intechopen.96257

Deol JK, Bains K (2010) Effect of household cooking methods on nutritional and anti nutritional factors in green cowpea (*Vigna unguiculata*) pods. J Food Sci Technol 4(7):579–581

Dhanasekar P, Reddy KS (2015) In: Proceedings of the National Conference on Pulses: Challenges and Opportunities under Changing Climatic Scenario, September 20- October 1, 2014, ISPRD, JNKVV, Jabalpur, pp. 316–320

Dhanasekar P, Reddy KS (2017) Role of seed Raffinose family oligosaccharides and Phytic acid in better performance potential of cowpea under water stress conditions. J Basic Appl Plant Sci 1 (2017):106

Dhanasekar P, Souframanien J, Suprasanna P (2021) Breeding cowpea for quality traits: A genetic biofortification perspective. In: Gupta DS, Gupta S, Kumar J (eds) Breeding for enhanced nutrition and bio-active compounds in food legumes. Springer, Cham, pp 157–179. https://doi.org/10.1007/978-3-030-59215-8_7

Dhole VJ, Reddy KS (2016) Association of phytic acid content with biotic stress tolerance in mungbean (*Vigna radiata* L. Wilczek). Phytoparasitica 44:261–267

Doblado R, Frias J, Vidal-Valverde C (2007) Changes in vitamin C content and antioxidant capacity of raw and germinated cowpea (*Vigna sinensis* var. *carilla)* seeds induced by high pressure treatment. Food Chem 101:918–923

Dugje IY, Omoigui LO, Ekeleme F, Kamara AY, Ajeigbe H (2022) Farmers' Guide to Cowpea Production in West Africa; IITA: Ibadan, Nigeria, 2009; Available online: http://www.icrisat.org/tropicallegumesII/pdfs/Cowpea.pdf

EDITS-Cowpea (2022). https://www.jircas.go.jp/en/database/edits-cowpea/introduction. Accessed on 07-08-2022

Eshwarage IS, Herath T, Gunathilake T (2017) Dietary fibre, resistant starch and in-vitro starch digestibility of selected eleven commonly consumed legumes (mung bean, cowpea, soybean and horse gram) in Sri Lanka. Res J Chem Sci 7:27–33

Fabbri ADT, Crosby GA (2016) A review of the impact of preparation and cooking on the nutritional quality of vegetables and legumes. Int J Gastron Food Sci 3:2–11

FAO (2017) The future of food and agriculture–trends and challenges. FAO, Rome

FAO (2020) World Food and Agriculture - Statistical Yearbook 2020, Rome

Felicio MR, Silva ON, Goncalves S, Santos NC, Franco OL (2017) Peptides with dual antimicrobial and anticancer activities. Front Chem 5:5

Fernandes Santos CA, Boiteux LS, Fernandes Santos CA et al (2015) Genetic control and transgressive segregation of zinc, iron, potassium, phosphorus, calcium, and sodium accumulation in cowpea (*Vigna unguiculata*) seeds. Genet Mol Res 14(1):259–268

Flores T, Karpova O, Su X et al (2008) Silencing of GmFAD3 gene by siRNA leads to low alpha-linolenic acids (18:3) of fad3-mutant phenotype in soybean [Glycine max (Merr.)]. Transgenic Res 17:839–850

Fróna D, Szenderák J, Harangi-Rákos M (2019) The challenge of feeding the world. Sustainability 11(20):5816. https://doi.org/10.3390/su11205816

Frota Kde M, dos Santos Filho RD, Ribeiro VQ, Arêas JA. (2015) Cowpea protein reduces LDL-cholesterol and apolipoprotein B concentrations, but does not improve biomarkers of inflammation or endothelial dysfunction in adults with moderate hypercholesterolemia. Nutr Hosp 31 (4):1611–9

Gerrano AS, van Rensburg WSJ, Ventera SL, Shargie NG, Amelework BA, Shimelis HA, Labuschagne MT (2019) Selection of cowpea genotypes based on grain mineral and total protein content. Acta Agric Scand Sect B Soil Plant Sci 69:155–166

Gerrano AS, Mbuma NW, Mumm RH (2022) Expression of nutritional traits in vegetable cowpea grown under various South African agro-ecological conditions. Plan Theory 11:1422

Gonçalves A, Goufo P, Barros A, Domínguez-Perles R, Trindade H, Rosa EA, Ferreira L, Rodrigues M (2016) Cowpea (*Vigna unguiculata* L. Walp), a renewed multipurpose crop for a more sustainable agri-food system: nutritional advantages and constraints. J Sci Food Agric 96(9): 2941–2951

González AM, Yuste-Lisbona FJ, Saburido S, Bretones S, De Ron AM, Lozano R et al (2016) Major contribution of flowering time and vegetative growth to plant production in common bean as deduced from a comparative genetic mapping. Front Plant Sci 7:1940

Gutierrez-Uribe JA, Romo-Lopez I, Serna-Saldivar SO (2011) Phenolic composition and mammary cancer cell inhibition of extracts of whole cowpeas (*Vigna unguiculata)* and its anatomical parts. J Funct Foods 3:290–297

Ha TJ, Lee M-H, Jeong YN, Lee HJ, Han S, Park HC et al (2010) Anthocyanins in cowpea [*Vigna unguiculata* (L.)Walp. ssp. *unguiculata*]. Food Sci Biotechnol 19:821–826

Hachibamba T, Dykes L, Awika J, Minnaar A, Duodu KG (2013) Effect of simulated gastrointestinal digestion on phenolic composition and antioxidant capacity of cooked cowpea (*Vigna unguiculata*) varieties. Int J Food Sci Technol 48:2638–2649

Herniter IA, Muñoz-Amatriaín M, Close TJ (2020) Genetic, textual, and archaeological evidence of the historical global spread of cowpea (*Vigna unguiculata* [L.] Walp.). Legume Sci 2:40–57

Horn L, Shimelis H, Sarsu F et al (2017) Genotype-by-environment interaction for grain yield among novel cowpea (*Vigna unguiculata* L.) selections derived by gamma irradiation. Crop J 6 (3):306–313

Iqbal A, Khalil IA, Ateeq N, Khan MS (2006) Nutritional quality of important food legumes. Food Chem 97:331–335

Jayathilake C, Visvanathan R, Deen A, Bangamuwage R, Jayawardana BC, Nammi S, Liyanage R (2018) Cowpea: an overview on its nutritional facts and health benefits running title: nutritional and health properties of cowpea. J Sci Food Agric 98:4793–4806

Ji J, Zhang C, Sun Z, Wang L, Duanmu D, Fan Q (2019) Genome editing in cowpea *Vigna unguiculata* using CRISPR-Cas9. Int J Mol Sci 20(10):2471

Kamara AY, Ewansiha SU, Ajeigbe HA, Okechukwu R, Tefera H, Boukar O, Omoigui LO (2010) Improvements in grain and fodder yield of cowpea (*Vigna unguiculata*) varieties developed in The Sudan savannas of Nigeria over the past four decades. In: Proceedings of the FifthWorld Cowpea Conference, Saly, Senegal, pp. 171–188

Kanetro B (2015) Hypocholesterolemic properties of protein isolate from cowpeas (*Vigna unguiculata*) sprout in normal and diabetic rats. Procedia Food Sci 3:112–118

Kapravelou G, Martínez R, Andrade AM, Chaves CL, López-Jurado M, Aranda P, Arrebola F, Cañizares FJ, Galisteo M, Porres JM (2015) Improvement of the antioxidant and hypolipidaemic effects of cowpea flours (*Vigna unguiculata*) by fermentation: results of in vitro and in vivo experiments. J Sci Food Agric 95:1207–1216

Khalid II, Elharadallou SB (2012) Functional properties of cowpea (*Vigna ungiculata* L. Walp), and Lupin (*Lupinus termis*) flour and protein isolates. J Nutr Food Sci 3:1–6

Kongjaimun A, Somta P, Tomooka N, Kaga A, Vaughan DA, Srinives P (2013) QTL mapping of pod tenderness and total soluble solid in yardlong bean [*Vigna unguiculata* (L.) Walp. subsp. *unguiculata* cv.-*gr. sesquipedalis*]. EUPHYTICA. 189(2):217–223

Laila O, Murtaza (2014) Seed sprouting: a way to health promoting treasure. Int J Curr Res Rev 6: 70–74

Lattanzio V, Terzano R, Cicco N, Cardinali A, Di Venere D, Linsalata V (2005) Seed coat tannins and bruchid resistance in stored cowpea seeds. J Sci Food Agric 85:839–846

Lazaridi E, Ntatsi G, Savvas D, Bebeli PJ (2016) Diversity in cowpea (*Vigna unguiculata* (L.) Walp.) local populations from Greece. Genet Resour Crop Evol 64:1529–1551

Liyanage R, Perera OS, Wethasinghe P, Jayawardana BC, Vidanaarachchi JK, Sivaganesan R (2014) Nutritional properties and antioxidant content of commonly consumed cowpea cultivars in Sri Lanka. J Food Legum Indian J Pulses Res 27:215–217

Lonardi S, Muñoz-Amatriaín M, Liang Q, Shu S, Wanamaker SI, Lo S, Tanskanen J, Schulman AH, Zhu T, Luo M-C, Alhakami H, Ounit R, Hasan AM, Verdier J, Roberts PA, Santos JR, Ndeve A, Doležel J, Vrána J, Hokin SA, Farmer AD, Cannon SB, Close TJ (2019) The genome of cowpea (*Vigna unguiculata* [L.] Walp.). Plant J 98:767–782

Mahalakshmi V, Ng Q, Lawson M, Ortiz R (2007) Cowpea [*Vigna unguiculata* (L.)Walp.] core collection defined by geographical, agronomical and botanical descriptor. Plant Genet Resour Charact Util 5:113–119

Marques MR, Soares Freitas RAM, Correa CAC, Siguemoto ES, Fontanari GG, Areas JAG (2015) Peptides from cowpea present antioxidant activity, inhibit cholesterol synthesis and its solubilisation into micelles. Food Chem 168:288–293

Marques MR, Maureira ADC, Fontanari GG, Pimenta DC, Soares-Freitas RM, Hirata HM et al (2018) Transport of cowpea bean derived peptides and their modulator effects on mRNA expression of cholesterol-related genes in Caco-2 and HepG2 cells. Food Res Int 107:165–171

Mekonnen TW, Gerrano AS, Mbuma NW, Labuschagne MT (2022) Breeding of vegetable cowpea for nutrition and climate resilience in sub-Saharan Africa: Progress, opportunities, and challenges. Plan Theory 11(12):1583. https://doi.org/10.3390/plants11121583

Mir RR, Kumar A, Pandey MK, Isobe SN (2021) Editorial: achieving nutritional security and food safety through genomics-based breeding of crops. Front Nutr 8:638845

Mohammed SB, Dzidzienyo DK, Umar ML, Ishiyaku MF, Tongoona PB, Gracen V (2021) Appraisal of cowpea cropping systems and farmers' perceptions of production constraints and preferences in the dry savannah areas of Nigeria. CABI Agric Biosci 2:25

Monteiro Junior JE, Valadares NF, Pereira HD, Dyszy FH, Da Costa Filho AJ, Uchoa AF et al (2017) Expression in *Escherichia coli* of cysteine protease inhibitors from cowpea (*Vigna unguiculata*): the crystal structure of a single-domain cystatin gives insights on its thermal and pH stability. Int J Biol Macromol 102:29–41

Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S, Pottorff M et al (2009) A consensus genetic map of cowpea [*Vigna unguiculata* (L.) Walp.] and synteny based on EST-derived SNPs. Proc Natl Acad Sci U S A 106:18159–18164

Muñoz-Amatriaín M, Mirebrahim H, Xu P, Wanamaker SI, Luo M, Alhakami H, Alpert M, Atokple I, Batieno BJ, Boukar O, Bozdag S, Cisse N, Drabo I, Ehlers JD, Farmer A, Fatokun C, Gu YQ, Guo Y-N, Huynh B-L, Jackson SA, Kusi F, Lawley CT, Lucas MR, Ma Y, Timko MP, Wu J, You F, Barkley NA, Roberts PA, Lonardi S, Close TJ (2017) Genome resources for climate-resilient cowpea, an essential crop for food security. Plant J 89:1042–1054

Nedumaran S, Abinaya P, Jyosthnaa P, Shraavya B, Rao P, Bantilan C (2015) Grain Legumes Production, Consumption and Trade Trends in Developing Countries, Working Paper Series 60. ICRISAT, Patancheru, pp 1–64

Oboh HA, Agu K (2010) The effects of various traditional processing methods on the glycemic index and glycemic load of cowpeas (vigna unguiculata). Journal of Food Biochemistry 34:1332–1342

Odeigah PGC, Osanyinpeju AO, Myers GO (1998) Induced mutations in cowpea, *Vigna unguiculata* (Leguminosae). Rev Biol Trop, San José 46(3)

Ojwang LO, Yang L, Dykes L, Awika J (2013) Proanthocyanidin profile of cowpea (*Vigna unguiculata*) reveals catechin-O-glucoside as the dominant compound. Food Chem 139:35–43

Ojwang LO, Banerjee N, Noratto GD, Angel-Morales G, Hachibamba T, Awika JM, Mertens-Talcott SU (2015) Polyphenolic extracts from cowpea (Vigna unguiculata) protect colonic myofibroblasts (CCD18Co cells) from lipopolysaccharide (LPS)-induced inflammation–modulation of microRNA 126. Food Funct 6(1):146–54

Onyeka EU (2007) Glycemic and physiochemical properties of five common cowpea (*Vigna unguiculata*) cultivars in Nigeria. J Food Process Preserv 31:618–631

Owade JO, Abong G, Okoth M, Mwang'ombe AW (2020) A review of the contribution of cowpea leaves to food and nutrition security in East Africa. Food Sci Nutr 8:36–47

Oyewale RO, Bamaiyi LJ (2013) Management of cowpea insect pests. Sch Acad J Biosci Sch Acad J Biosci 1:217–226

Padhi SR, Bartwal A, John R, Tripathi K, Gupta K, Wankhede DP, Mishra GP, Kumar S, Archak S, Bhardwaj R (2022) Evaluation and multivariate analysis of cowpea [*Vigna unguiculata* (L.) Walp] germplasm for selected nutrients—mining for nutri-dense accessions. Front Sustain Food Syst 6:888041. https://doi.org/10.3389/fsufs.2022.888041

Pandey AK, Madhu P, Bhat BV (2019) Down-regulation of CYP79A1 gene through antisense approach reduced the cyanogenic glycoside Dhurrin in [Sorghum bicolor (L.) Moench] to improve fodder quality. Front Nutr 6:122

Pasquet RS, Feleke Y, Gepts P (2021) Cowpea [Vigna unguiculata (L.) Walp.] maternal lineages, chloroplast captures, and wild cowpea evolution. Genet Resour Crop Evol

Pathirana R, Carimi F (2022) Management and utilization of plant genetic resources for a sustainable agriculture. Plan Theory 11:2038. https://doi.org/10.3390/plants11152038
Perera O, Liyanage R, Jayawardana BC, Vidanarachchi JK, Fernando P, Sivaganesan R et al (2016) Modulating effects of cowpea incorporated diets on serum lipids and serum antioxidant activity in Wistar rats. J Natl Sci Found Sri Lanka 44:69
Popova A, Mihaylova D (2019) Antinutrients in plant-based foods: A review. Open Biotechnol J 13:68–76
Punniyamoorthy D, Reddy KS, Dhanasekar SP (2007) IANCAS bulletin, Nov 2007, pp. 299–307
Raina A, Laskar RA, Wani MR, Jan BL, Ali S, Khan S (2022) Gamma rays and sodium Azide induced genetic variability in high-yielding and biofortified mutant lines in cowpea [*Vigna unguiculata* (L.) Walp.]. Front Plant Sci 13:911049. https://doi.org/10.3389/fpls.2022.911049
Santos CAF, Boiteux LS (2013) Breeding biofortified cowpea lines for semi-arid tropical areas by combining higher seed protein and mineral levels. Genet Mol Res 12:6782–6789
Santos CAF, da Costa DCC, da Silva WR, Boiteux LS (2012) Genetic analysis of total seed protein content in two cowpea crosses. Crop Sci 52:2501–2506
Tadele Z (2019) Orphan crops: their importance and the urgency of improvement. Planta 250:677–694
Talabi AO, Vikram P, Thushar S, Rahman H, Ahmadzai H, Nhamo N, Shahid M, Singh RK (2022) Orphan crops: A best fit for dietary enrichment and diversification in highly deteriorated marginal environments. Front Plant Sci 13:839704
The Kirkhouse Trust. http://www.kirkhousetrust.org/
Trehan I, Benzoni NS, Wang AZ, Bollinger LB, Ngoma TN, Chimimba UK et al (2015) Common beans and cowpeas as complementary foods to reduce environmental enteric dysfunction and stunting in Malawian children: study protocol for two randomized controlled trials. Trials 16:520
Udenigwe CC, Rouvinen-Watt K (2015) The role of food peptides in lipid metabolism during dyslipidemia and associated health conditions. Int J Mol Sci 16:9303–9313
Wu X, Islam ASMF, Limpot N, Mackasmiel L, Mierzwa J, Cortés AJ, Blair MW (2020) Genome-wide SNP identification and association mapping for seed mineral concentration in mung bean (Vigna radiata L.). Front Genet 11:656
Xiong S, Yao X, Li A (2013) Antioxidant properties of peptide from cowpea seed. Int J Food Prop 16:1245–1256

Lentils (*Lens culinaris* Medik): Nutritional Profile and Biofortification Prospects

Debjyoti Sen Gupta, Jitendra Kumar, Surendra Barpate, A. K. Parihar, Anup Chandra, Anirban Roy, and Ivica Djalovic

Contents

D. Sen Gupta (✉) · J. Kumar · A. K. Parihar · A. Chandra
ICAR-Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India
e-mail: debjyoti.gupta@icar.gov.in

S. Barpate
Scientist, Food Legumes Research Platform (FLRP), International Centre for Agricultural Research in the Dry Areas (ICARDA), Sehore, Madhya Pradesh, India

A. Roy
Ramakrishna Mission Vivekananda Educational and Research Institute, Kolkata, West Bengal, India

I. Djalovic
Institute of Field and Vegetable Crops, National Institute of the Republic of Serbia, Novi Sad, Serbia

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_27

Abstract

Lentil is a highly consumed pulse crop in India, Bangladesh, Nepal, and many other countries. This crop is rich in nutrients which are easy to digest and palatable. In the form of a whole food source of nutrition, it can minimize the effect of malnutrition which is prevalent worldwide. Biofortification as a tool provides us great opportunity to further enhance nutrient content biologically. A few studies showed considerable genetic variability for nutrients including iron, zinc, calcium, magnesium, selenium, prebiotic carbohydrates, and folate concentration in lentil, which may be further improved. While breeding for nutrients, the role of environmental effects should be taken into consideration to provide a widely adapted plant variety. A number of genomic regions have been mapped using molecular markers; however, the intensity and coverage of the experiments were low, and this area needs more efforts to make marker-assisted breeding a reality in lentil breeding for nutritional traits. Lentil also has anti-nutrients like phytic acid, which influences the bioavailability of nutrients. In addition to traditional breeding approaches, efforts are underway to make use of cis- or transgenic technologies to enhance nutritional quality in many crops; the same may be adopted based on need in the case of lentil. Biofortified lentil varieties recently released which are rich in iron and zinc concentration; however, more varieties are required to cover different agroclimatic regions or niches. In short, more focused efforts are required to identify high-yielding, biotic and abiotic stress-tolerant, and nutrient-rich new-generation lentil varieties that will definitely boost health status among the consumers, especially from today's perspective when plant-based protein or other nutrients are gaining huge popularity.

Keywords

Lens culinaris · Micronutrient · Biofortification · Bioavailability · Anti-nutrients · Iron · Zinc

1 Introduction

Malnutrition is caused by unavailability of sufficient quantity of nutrients like proteins, carbohydrates, micronutrients, and vitamins. Higher quantity of anti-nutrients in our everyday meals is also detrimental to our health. This situation is present in both developed and developing countries. Consumption of carbohydrate-rich diets throughout the world aggravated the situation further since these are deficient in highly essential micronutrients, causing long-lasting disorders or diseases among the people (Stewart et al. 2010; Bouis et al. 2011). Thus, "hidden hunger" shows its effects on the well-being of the micronutrient-starved population, though in many cases have sufficient calorie intake. Implication of such long exposure to low or less than recommended requirement of micronutrients or vitamins causes serious health issues including lower birth weight, anemic condition, impaired learning ability,

high mortality and morbidity, reduced working ability, and increased treatment costs (Batra and Seth 2002; Welch and Graham 1999). Deficiencies of iron (Fe), zinc (Zn), selenium (Se), and iodine (I) are not uncommon in South and South-East Asian countries along with Indian subcontinent. According to an estimate, nearly 60% of the global population is usually facing deficit of Fe (Yang et al. 2007) whereas 33% of Zn (Hotz et al. 2004) and almost 15% of Se (FAOSTAT 2007).

About one-fourth of the people globally complain of anemia (WHO 2008), whereas 17.3% people are devoid of sufficient Zn intake (Wessells and Brown 2012), causing 433,000 deaths annually of children below age 5 (WHO 2009). About 20% children morbidity of age less than 5 years were due to the inadequate intake of vitamins like vitamin A, Zn, Fe, and/or I (Prentice et al. 2008). In preschool children and pregnant and lactating women, deficiencies of essential micronutrients especially Zn and Fe are observed more common (Welch and Graham 1999; White and Broadley 2009; WHO 2012). About one billion people suffered with Keshan (cardiomyopathy) and Kashin–Beck (osteoarthropathy) diseases caused by Se deficiency (Reilly 1996). Countries like Australia, New Zealand, Africa, the UK, Thailand, Finland, Central Siberia, Denmark, Turkey, Northeast to South-Central China, parts of India, Bangladesh, and Nepal have low levels of soil Se which is bioavailable (100–2000 $\mu g\ kg^{-1}$), and due to this, crops grown in these countries have low concentration of Se in them, creating a perfect ground for Se deficiency-related health issues (Fordyce 2005; Lyons et al. 2005; Spallholz et al. 2004, 2008). Besides the role of a typical micronutrient, Se prevents cytotoxic effect of arsenic (Biswas et al. 1999); arsenic toxicity exists in many crop ecologies in South Asia, as Se and As detoxify each other's effect (Holmberg Jr and Ferm 1969; Levander 1977). Similarly, deficiency of folate prevails worldwide over millions and leads to serious health problems, including birth defects and health risks both for the mother and child (Gupta et al. 2013).

To combat the global malnutrition-related issues, several mitigations have been adopted by governments, nongovernmental organizations, and the United Nation. Food fortification through supplements is effective to bolster health in many communities from both developed and developing world. In fact, for many processed or polished food grains or products, food fortification is mandatory in many countries. However, food fortification causes price of the fortified food to move upward beyond the capacity of common consumers in target communities who are poor or have other priorities than to spend higher for fortified food purchase for consumption. Here comes the role of another very unique and successful strategy to develop biofortified crop varieties, which usually requires one-time investment (Bouis et al. 2011).

Food crops can be biofortified in two ways: one of the popular method is to increase the concentration of target micronutrient by the external application of such micronutrient over the plant part for their ready absorption followed by translocation within the plant system and ultimately in the grain or any other plant part which is consumed. Crop rotations and intercropping including soil microbes may also be the components in agronomic biofortification helping in transport of target micronutrient in the soil (White and Broadley 2009). There are many success stories available for agronomic biofortification like Se can be achieved by its spraying in lentil

(Thavarajah et al. 2015) and so in crops like potato tubers, field pea seeds, and tea leaves (Smrkolj et al. 2006; Turakainen 2007). Increase in production cost, particularly in developing countries, agronomic biofortification is not always feasible among the growers particularly for the smaller-land holdings (Graham and Rengel 1993). Complexity in the application of perfect crop growth stages as this method is vulnerable to both external and internal crop growing conditions (Terry et al. 2000). Application concentration of micronutrients and any allied inorganic nutrients should be standardized in many crops, thereby reducing the chances of any induced phytotoxicity. Secondly, genetic biofortification is a plant breeding-based technique wherein nutritionally rich high yielding food crop varieties are achieved by manipulating the genetic buildup of cultivars using either classical plant breeding tools or modern genomic approaches. This approach offers an one-time investment and cost-effective method of delivering essential micronutrients and, thereby, does not require repeated efforts to fortify during food processing; therefore, biofortified seeds can be supplied to the target community who can easily grow, produce, and maintain the seed chain for a long period (Graham et al. 2007; White and Broadley 2009). It is a tool for food crop improvement, and sustenance which have been discussed by Miller and Welch (2013) and Saltzman et al. (2013). Several workers have discussed approaches to enhance zinc, selenium, and iron (Hawkesford and Zhao 2007; Velu et al. 2014) and deployment of genomics for biofortification of Fe and Zn in wheat (Borrill et al. 2014) and in common bean (Blair 2013; Petry et al. 2015). Genetic potential for high Fe and Zn concentrations and low anti-nutrients like PA in various food crops such as maize (*Zea mays* L.), rice (*Oryza sativa* L.), wheat (*Triticum sativum* L.), common bean (*Phaseolus vulgaris* L.), and field pea had been reviewed in recent years (Frossard et al. 2000; Gomez-Galera et al. 2010; Amarakoon et al. 2012). This chapter will provide information on the genetic potential of lentil as a whole food biofortified crop and recent achievements made in this direction.

2 Description on Nutritional Components

Lentil (*Lens culinaris* subsp. *culinaris* Medikus) is a nutritious food legume crop. It is cultivated in more than 50 countries but majorly grown in India (36%), western Canada (18%), south-eastern Turkey (15%), and Australia (4%) (FAOSTAT 2011). Like other pulses, it is rich in proteins, zinc, iron, folate, selenium, vitamins, and carotenoids (Thavarajah et al. 2011a, b; Johnson et al. 2013a, b; Gupta et al. 2013) (Table 1). Serving 100 g of lentil grain can fulfill the recommended daily allowance (RDA) – Fe: 41–113%; Zn: 40–68%; and Se 77–122% (Table 1). The pulse is rich in ß-carotene (2–12 µg/g) but contains less phytic acid phosphorous (0.7–1.2 mg/g). The amount of phytic acid has been found relatively lower than cereal and pulse crops with low PA (rice: 2.23 mg/g; soybean: 4.86 mg/g; wheat: 2.51 mg/g; maize: 3.7 mg/g; and common bean: 1.38 mg/g). As a whole food, lentil can be cooked within 10 min which is a great time and energy saver (Thavarajah et al. 2011a). Therefore, lentil is a natural candidate for biofortification efforts to develop a nutritionally rich, high protein, high iron and zinc containing pulse crop which if

Table 1 Nutritional value of lentil seeds

Protein	20–25%
Carbohydrate	50–60%
Fat	0.7–0.8%
Ca	60–70 mg/100 g
Fe	7–9 mg/100 g
Zn	4–5 mg/100 g
Se	42–67 μg/100 g
Folate	261–290 μg/100 g

Source: Adapted from Kumar et al. (2016a)

eaten with cereals can fulfill both calorific as well as micronutrient requirements. This will further help to fight global malnutrition problem among poor populations or communities. The various nutritional/antinutritional compounds estimated in lentil grain are given in Table 1.

3 Traits Required for Development of Biofortified Lentil

Modern breeding approaches along with conventional tools were found to be highly suitable for the breeding of nutrient-dense crop cultivars with high bioavailability of phytonutrients (Nestel et al. 2006). To achieve it, first targets should be fixed with respect to the status of nutrients and anti-nutrients present in the grain or seed. After knowing this information, breeder can plan to improve the bioavailability of a nutrient either by increasing its concentration or by reducing the concentration anti-nutrients present. Main objective of major biofortification efforts in cereals and pulses involves high iron and zinc concentration as target for breeding. Beside this, an anti-nutrient (phytic acid), which reduces the bioavailability of iron and zinc, is also targeted for reduction of their concentration by introgressing the genes or quantitative trait loci (QTLs) that produce anti-nutritional phytochemicals in lower concentration in plants. These potential biofortification-related traits are presented in Table 2.

4 Genetic Resources of Health-Related (HR) Genes

A number of genebanks store cultivated and wild lentils in their repository (Table 3). The largest *Lens* collection (31,970) is Genesys followed by ICARDA genebank which conserved 13,958 *Lens* accessions in Morocco (Dikshit et al. 2022). Australia, Iran, the USA, Russian Federation, and India have 1000–6000 accessions in their respective genebanks (Table 3). Canada has 1139 *Lens* accessions in their genebank. Variable numbers of lentils (678–1095) are collected by Turkey Syria, Hungary, Egypt, China, Pakistan, Bangladesh, and Ethiopia for conservation in respective genebanks. Core and mini-core sets from local collection have been developed by many research groups which are valuable genetic resources for their use in breeding

Table 2 Potential areas for biofortification in lentil

Traits	Range
Nutritional traits	
Protein	15.9–32%
Starch	34.7–65.0%
Dietary fibers	5.1–26.6%
Fatty acids	0.3–3.5 g/100 g
Micronutrients	
Iron	73–90 mg/kg
Zinc	44–54 mg/kg
Sorbitol	1250–1824 mg/100 g
Mannitol	57–132 mg/100 g
Galactinol	46–89 mg/100 g
Sucrose	1750–2355 mg/100 g
Raffinose + stachyose	3314–4802 mg/100 g
Verbascose	1907–2453 mg/100 g
Nystose	8–450 mg/100 g
Anti-nutritional traits	
Phenolics	6.24–27.73 mg GAE/g defatted sample
Flavonoids	1.15–4.94 mg CE/g defatted sample
Condensed tannin content	3.14–12.97 mg CE/g defatted sample
Phytoestrogens	8.9–12.3 μg/100 g dry matter
Phytate	3.9–11.9 mg/g
Saponins	0.07–0.13 g/100 g
Protease inhibitor	25–55 TIA/mg of protein
α-amylase inhibitor	–
Lectins	–
Vicilin protein	–

Source: Modified from Kumar et al. (2016b)
GAE gallic acid equivalent, *CE* Catechin equivalent, *TIA* trypsin inhibitor activity

programs (Tripathi et al. 2022). However, there is a need in lentil for evaluation of these core or mini-core sets for nutrition-related traits.

ICARDA *Lens* accessions around 76% have been characterized for morphological and phenological attributes (Kumar et al. 2016c). Apart from this, under the AGILE project that is led by Canadian Scientists, 324 accessions of lentil diversity panel were phenotyped for phenological traits (different nine diverse locations for two seasons) around the world (Wright et al. 2021). This set has also been genotyped using an exome capture array and seeds from 321 lines have been deposited in the ICARDA genebank (Ogutcen et al. 2018).

Lentil is dense with protein contents and an affordable pulse crop that is an ancient crop and has a long domestication history. The lentil or *Lens* species belongs to the family Fabaceae and its crop wild relatives (CWR) are naturally and widely distributed in Mediterranean regions and Southwest Asia (Guerra-García et al. 2021). The genus *Lens* ($2n = 2x = 14$) phylogenetically belongs to the tribe Vicieae.

Table 3 Prominent genebanks and their holding size in lentil

Country	Name of center/genebank	Total *Lens* accessions including CWR conserved in different global genebanks	References
Global	Genesys	31,970	https://www.genesys-pgr.org
Morocco	ICARDA	13,958	Dikshit et al. 2022
Australia	Australia Temperate Field Crops Collection	5254	Singh et al. 2017; Dikshit et al. 2022
Iran	Seed and Plant Improvement Institute	3000	
USA	USDA	2875	
Russian Federation	N.I. Vavilov All-Russian Scientific Research Institute of Plant Industry	2556	
Chile	Inst de Inv. Agropecuaries, Centro Regional de Investigation Carillanca	1345	
Canada	PGRC	1139	
Syria	General Commission for Scientific Agricultural Research	1072	
Egypt	National Gene Bank	875	
Pakistan	Plant Genetic Resources Institute	805	
Bangladesh	BARI	798	
Spain	Centro de Recuros Fitogenetico, INIA	703	
Ethiopia	Biodiversity Conservation and Research Institute	678	
India	NBPGR, New Delhi	2285	
Hungary	Research Center for Agrobotany	1061	
China	Institute of Crop Germplasm Resources	855	
Turkey	Plant Genetic Resources Department, Aegean Agricultural Research Inst.	1095	

The tribe contained cool season legume crops which are members of the family Fabaceae (Ladizinsky 1979). Presently, the *Lens* genus contained seven closely related species mainly *Lens culinaris, L. orientalis, L. tomentosus, L. lamottei, L. odemensis, L. ervoides,* and *L. nigricans* (Table 4). The studied taxonomic relationships on the basis of morphology, hybridization behavior, cytogenetics, and molecular markers do not accept classification at the level of species and

Table 4 *Lens* gene pool and species

S. No	Gene pool	Species name	Type	References
1	Primary	*Lens culinaris*	Domesticated	Ladizinsky 1979; Wong et al. 2015; Guerra-García et al. 2021
		Lens orientalis	Wild	
		Lens tomentosus	Wild	
2	Secondary	*Lens odemensis*	Wild	
		Lens lamottei	Wild	
3	Tertiary	*Lens ervoides*	Wild	
4	Quaternary	*Lens nigricans*	Wild	

subspecies. Although, researchers commonly agree that *L. culinaris* ssp. *orientalis* is the wild progenitor of *L. culinaris* ssp. *culinaris* whereas *L. nigricans* is distantly related. Apart from this, *Lens culinaris* subsp. *culinaris* is subdivided into two types: the microsperma cultivars having small-sized seeds and another with reddish cotyledons which is believed to be originated from near East and Central Asia region and the macrosperma cultivars having large seeded red, yellow, and green cotyledon are native to Mediterranean region.

Based on a study (Schreier et al. 2012), genotyping data-based analysis of all the individuals like *L. culinaris/L. orientalis/L. tomentosus* and *L. ervoides/L. nigricans* each belong to one cluster whereas *L. lamottei and L. odemensis* exhibited diversified ancestry with *L. ervoides/L. nigricans cluster in* major proportion. The Chinese group of researchers (Wong et al. 2015) also categorized *Lens* in seven different taxa in four separate gene pool: as primary (*L. culinaris*), secondary (*L. orientalis*), tertiary (*L. tomentosus* and *L. lamottei)*, and quaternary (*L. ervoides* and *L. nigricans*). However, crossing incompatibility for hybridization in the genus *Lens* among the gene pools reported by Cubero et al. (2009). Many studies reported the crossability of *L. odemensis* and *L. orientalis* with *L. culinaris* (Fratini and Ruiz 2006; Muehlbauer et al. 2006). Whereas, some of the studies showed hybrid embryo abortion, albino seedlings, and hybrid sterility are major restrictions in the wide hybridization among *Lens* species. Gupta and Sharma (2006) stated that hybrid embryos abort while crossing the parents from the tertiary gene pool with *L. culinaris.* These hybridization barriers can be overcome through tissue culture-based embryo rescue methods (Tullu et al. 2013). Apart from this, application of GA3 (gibberellic acid) also impacted positively and produced successful crosses of *L. culinaris* with *L. ervoides, L. odemensis,* and *L. nigricans* (Fratini and Ruiz 2006).

In due course of lentil domestication process, uninterrupted selection reduced genetic variation/allelic diversity of members of primary gene pool in comparison to secondary, tertiary gene pool or wild progenitor. In the *Lens* primary gene pool,

reduced allelic variation during domestication is the major hindrance for enhancing productivity and other useful agronomical traits. Because of the impact of climate change and unpredictable weather condition during growing season, effective adaptation strategies to combat with biotic and abiotic stresses are required.

Therefore, crop wild relatives (CWRs) have remarkable potential and reservoir of economically important genes/alleles to enhance crop productivity as well as the tolerance against the biotic and abiotic stresses. Hence, CWRs are required for lentil improvement as potential sources of useful gene(s) for the deployment in cultivars. Thus, wild relatives act as most important donors for added genetic variability in cultivated crops.

5 Genetic Variability for Biofortification Traits in *Lens* Gene Pool

5.1 Primary Gene Pool

Since the inception of the HarvestPlus program, for many crops, identification of natural variants with high concentrations of iron and zinc in the cultivated gene pool has been a principal approach for the development of biofortified crop cultivars. As the natural variants are identified, if they are of less agronomic value, then such nutritional traits are transferred to agronomic bases by either hybridization or other techniques. As discussed by Kumar et al. (2016a), screening for high iron and zinc concentration in lentil was conducted in many countries, including India, Turkey, Syria, Canada, and Pakistan (Table 5). In these studies, folate, iron, and zinc concentrations have been targeted. In a study involving 19 lentil genotypes pertaining to the concentration of Fe and Zn from various markets of Canada, narrow range of genetic variability was observed. Fe concentration was found ranging between 73 and 90 mg of Fe kg^{-1} whereas Zn concentration between 44 and 54 mg of Zn kg^{-1} in these genotypes (Thavarajah et al. 2009a). This was followed by a large-scale experiment conducted by the International Center for Agricultural Research in the Dry Areas (ICARDA) under the HarvestPlus Challenge program in which they have screened iron (42–132 ppm) and zinc (23–78 ppm) concentration in a large collection of more than 1500 lentil accessions from cultivated gene pool (Sarker et al. 2007; HarvestPlus 2014). In a multiyear and multilocation study with respect to concentration of folate, significant genetic variability was observed which ranged between 216 and 290 μg/100 g in 10 lentil cultivars of the USA. Compared to other pulses like chickpea, yellow field pea, and green field pea, lentil was having more folate concentration than other pulses (Gupta et al. 2013). Geographical areas with limited zinc availability in soils, including India, Pakistan, China, Iran, and Turkey, are also countries where human Zn deficiency is most prevalent (Hotz et al. 2004; Khan et al. 2008). Eyüpoğlu et al. found that more than 50% of the land (14 Mha) in Turkey is Zn deficient. The high prevalence of Zn-deficient soils in Turkey has been suggested as a major cause of Zn deficiency and to be indirectly related to deficiencies of other micronutrients. Turkish landraces of lentil also

Table 5 Genetic variability for biofortification traits in lentil

Type of genetic material	No. of accessions	Compound with range of concentration	Country	References
Landraces, wild types, and breeding lines	1600	Fe: 43–132	ICARDA, Syria	Sarker et al. 2007
		Zn: 22–78	do	Baum et al. 2008
Lentil germplasm		Fe: 41–109 Zn: 22–78		
Breeding lines, germplasm, and modern high-yielding genotypes	900	Fe: 73–90	Canada	Thavarajah et al. 2008, 2009a
		Zn: 44–54		
		Se: 425–673		
Landraces and cultivars	46	Fe: 49–81	Turkey	Karakoy et al. 2012
		Zn: 42–73		
Landraces and breeding material	96	Fe: 37–157		
		Zn: 26–65	India	Kumar et al. 2018a
Exotic lines		Se: 240–630		
Varieties	19	Fe: 73–90	Canada	Thavarajah et al. 2009a
		Zn: 44–54 mg		
Varieties		Folate: 216–290 μg/100 g	The USA	Gupta et al. 2013
Elite breeding lines	41	Fe: 50.85–136.9 mg	India	Kumar et al. 2014a
		Zn: 40.26–81.5		
Varieties	7	Se: 74–965 μg kg^{-1}	Bangladesh	Rahman et al. 2013
Varieties	23	Fe: 43–92 ppm	Canada	DellaValle et al. 2013
		PA: 3.8–15.9 mol/g		
Germplasm line	–	Beta-carotene: 2–12 μg/g	The USA	Thavarajah et al. 2011b
Varieties	9	Prebiotic carbohydrate	The USA	Johnson et al. 2013b
Cultivars and breeding lines	192	Se: 6–254 μg/kg^{-1}	Syria, Nepal, Morocco, the USA, Australia, and Turkey	Thavarajah et al. 2011b

exhibited a range of genetic diversity not only for different micronutrients (Karakoy et al. 2012).

In India, in a study, 41 elite lentil lines were tested for stability of grain Fe and Zn concentration across multiple locations. Pooled analysis of variance over locations detected significant differences between genotypes, locations, and genotype × location interaction. The average grain Fe concentration over the locations was obtained for L 4704 (137 mg/kg grain), while the grain Zn concentration was highest for VL 141 (82 mg/kg grain). Although both micronutrients were affected by the environment, iron concentration exhibited more G × E interaction compared to Zn concentration (Kumar et al. 2014a). Lentil seeds are also a very good source of organic Se (selenomethionine) (Thavarajah et al. 2007, 2008), and it was also reported that cooking had limited modifying effects on selenomethionine concentration (Thavarajah et al. 2008). In case of Se, significant genetic variability was noted in the genotypes from various countries (Thavarajah et al. 2011b; Rahman et al. 2013). In Bangladesh, total Se concentration was estimated in soil and lentil seeds received from both farmers' fields and yield trials. Average of Se concentration in farmers' fields was 163 and 312 μg kg^{-1}, for soil and lentil seed, respectively. It was further calculated that 50 g of lentil consumption contributes 28% of the recommended daily allowance of Se (55 μg per person per day). Relative bioavailability of Fe was estimated through in vitro digestion/Caco-2 cell model in 23 genotypes of cultivated gene pool of lentil, wherein significant genetic variability reported for Fe and relative iron bioavailability and phytic acid (PA) concentration (DellaValle et al. 2013). The pulse is also a great source of beta-carotene where significant genetic variability (2–12 μg/g) was reported in the USA (Thavarajah et al. 2011b). Prebiotic carbohydrates are important to maintain gut microflora, and significant genetic variation exits in lentil for this complex biochemical trait (Tahir et al. 2011; Wang et al. 2009). Johnson et al. (2013b) reported the genetic variability in prebiotic carbohydrates such as raffinose-family oligosaccharides, sugar alcohols, fructooligosaccharides, and resistant starches in 10 commercially grown lentil varieties of the USA from cultivated genepool, which may be further elevated through the approach of breeding by utilizing more number of germplasm or varieties under screening. Significant environmental effect on this type of traits demands site-specific breeding (Johnson et al. 2013b).

5.2 Secondary or Tertiary Gene Pool

Wild gene pool or alien gene pool or secondary and tertiary gene pool always are sources of lost useful genes or alleles, which were lost from the cultivated species in due course of evolution (Doyle 1988; Tanksley and McCouch 1997). The breeding history of a crop species has a bearing on the extent, type, and size of such wild sources of alleles. Being the most abundant curator of genetic resources in lentil,

ICARDA in Lebanon holds more than 500 wild lentil accessions, representing 6 wild *Lens* species from 26 countries. Many species of wild *Lens* genepool have expressed their cross compatibility with cultivated species (Fratini et al. 2004; Fratini and Ruiz 2006; Muehlbauer et al. 2006) after employing modifications in tissue culture-mediated hybridization procedures (embryo rescue) or by using different doses of plant growth regulators. Accessions from wild genepool have been identified as donors for high concentration of micronutrients in many crops (Cakmak et al. 2000; Ortiz-Monasterio et al. 2007).

Seed micronutrient concentration of cultivated and wild lentils is not much known. Sen Gupta et al. (2016) estimated micronutrients in the seeds of 26 lentil genotypes, representing 4 species and 3 subspecies of *Lens*. Concentrations of Fe, Zn, Ca, Cu, and Mg in seeds varied from 26 to 92, 17–51, 97–536, 3–12, and 272–892 mg kg^{-1}, respectively, among the *Lens culinaris* genotypes. Mineral concentrations for *L. lamottei* (Fe = 64–80, Zn = 26–40, Ca = 311–434, Cu = 2–6, and Mg = 754–839 mg kg^{-1}), *L. nigricans* (60–70, 33–39, 508–590, 3–4, and 445–738 mg kg^{-1}), and *L. ervoides* (65, 37, 339, 6, and 638 mg kg^{-1}) were similar to *Lens culinaris* genotypes. More number of germplasm should be tested in future to identify higher genetic variation in lentil for these micronutrients.

Kumar et al. (2018b) evaluated a core set of 96 wild accessions extracted from 405 global wild annual collections comprising different wild *Lens* species for micronutrients. Significant genetic variation was observed for different micronutrients including Na (30–318), K (138.29–1578), P (37.50–593.75), Ca (4.74–188.75), Mg (15–159), Fe (2.82–14.12), Zn (1.29–12.62), Cu (0.5–7.12), Mn (1.22–9.99), Mo (1.02–11.89), Ni (0.16–3.49), Pb (0.01–0.58), Cd (0–0.03), Co (0–0.63), and As (0–0.02) (mg/100 g). Significant positive correlations among micronutrients were also observed. It is noteworthy to mention that accessions representation from Turkey and Syria had maximum variability for different micronutrients. Hence, wild gene pool can be used to transfer favorable alleles controlling higher micronutrient concentrations in lentil.

6 Classical Genetics and Traditional Breeding of HR Traits

6.1 Genetics of HR Genes

Traits like iron and zinc concentration and folate concentration are quantitatively inherited. In cereals, it was reported long ago that these kinds of traits are controlled by several genes with individually smaller effects on the trait mean. In wheat, generation mean analysis was conducted to find out the genetics of iron and zinc concentration (Amiri et al. 2020). Due to the role of fixable gene effects and high heritability for grain iron concentration, selection for this trait could be effective in early generations. Grain zinc concentration was having nonadditive gene effects and low heritability; hence selection should be practiced during advanced generations. Both kind of heterosis at both crosses for iron and zinc concentration were significantly negative. Response to selection varied over locations for iron and zinc concentrations.

6.2 Site-Specific Breeding for Biofortification Traits

Environmental conditions during crop growth including soil pH, soil temperature, photoperiod, rainfall, soil organic matter, and texture affect the rate of accumulation of micronutrients into seeds (Cakmak 2008; Joshi et al. 2010). Hence, knowledge of such an optimal environment for growing facilitates the sink deposition of iron and zinc in seed, which is the key to the success of any biofortification research program. Here comes the role of genotype × environment interactions, which is usually very high for traits involving secondary metabolism. It is important to encourage breeders to develop crop cultivars that are site-specific and not look for wider adaptation, as that will be contrary to the overall breeding behavior of these traits.

In lentil, iron, zinc, and phytic acid concentrations vary over locations, soil conditions, and weather parameters (Thavarajah et al. 2009a, b; Kumar et al. 2018a). It was further observed that samples from one geographical region varied from another (Thavarajah et al. 2011a). For example, Fe concentration was higher in samples of Syria (63 mg/kg), Turkey (60 mg/kg), and the USA (56 mg/kg). Likewise, Zn concentration was higher in the seeds from Syria (36 mg/kg), Turkey (32 mg/kg), and the USA (28 mg/kg). Se was high in the seeds from Nepal (180 μg/kg) and Australia (148 μg/kg) to Syria (22 μg/kg), Morocco (28 μg/kg), and Turkey (47 μg/kg) (Thavarajah et al. 2011a). Calcium was high in samples from Turkey (0.48–1.28 g/kg), while lentils grown in India have high Fe (37–156 mg/kg) and Zn (26–65 mg/kg) concentrations (Karakoy et al. 2012; Kumar et al. 2016a).

A significant year × location interaction has been observed for total folate concentration in 10 lentil cultivars from the USA in a study conducted over 2 years (Gupta et al. 2013). It was observed that temperature influences the concentration of iron (Fe), phytic acid (PA), and zinc (Zn) in lentil genotypes on growing under different temperature regimes: Saskatoon, Canada (decreasing temperatures) and Lucknow, India (increasing temperatures).

In lentil seeds, concentrations of Fe, PA, and Zn were significantly higher in the regime with a rising temperature than in the regime with a decreasing temperature. Microclimatic factors control the iron, zinc, and phytic acid concentrations in lentil, which show a quite similar trend as compared to other candidate crops (Gupta et al. 2013). Under changing climatic conditions, fluctuating temperature regimes (particularly warm night temperature) may affect the anti-nutrients like phytic acid biosynthesis in a negative way, further complicating the biofortified varietal improvement in this crop species.

6.3 Breeding for Biofortified Lentil Cultivars

Initially, through the HarvestPlus program, several already released varieties were tested for iron and zinc concentration (i.e., Ethiopia, Bangladesh, Morocco, Turkey, Syria, and Nepal). Many of these varieties which were identified as high yielding as well as having high iron and zinc concentration are listed in Table 3. A few of these lines are promoted by the national partners to reach to lentil growers rapidly. In

Bangladesh, high iron and zinc containing varieties such as Barimasur 4 and Barimasur 5 are promoted to all corners of the country.

HarvestPlus (2014) reported that several released varieties of lentil that possess high iron and zinc concentrations and high yield potential have been identified in Nepal; they are: Sisir (98 ppm Fe and 64 ppm Zn), Khajurah-2 (100.7 ppm Fe and 59 ppm Zn), Khajurah-1 (58 ppm Zn), Sital (59 ppm Zn), Shekhar (83.4 ppm Fe), and Simal (81.6 ppm Fe). In Nepal, another lentil variety, ILL 7723 (Khajura-4), was released for general cultivation in Nepal (Darai et al. 2020) with high iron and zinc concentration. This is a selection in materials from ICARDA, Morocco (Sel89503). In India, two biofortified lentil varieties, Pusa Ageti Masoor (Fe 65 ppm) and IPL 220 with high Fe (73–114 ppm) and Zn (51–65 ppm) concentrations, have been released so far and both of these became very popular among the lentil growers. Pusa Ageti Masoor is covering the central part and IPL 220 covers the northeastern plain zone of the country.

7 Genetic Analysis of Agronomic Traits

The *Lens* gene pool, especially *Lens culinaris* (lentil), and its wild progenitor are successfully cultivated in areas of low rainfall (less than 250 mm/year) on marginal and poor soil. Available germplasm in the primary gene pool demonstrated poor water extraction capacity and a lower growth rate as compared with secondary and tertiary gene pool. It is well known that the most prominent abiotic stress is drought.

The drought tolerance capacity has been identified and reported in *L. ervoides, L. nigricans, and L. odemensis* (Gupta and Sharma 2006). The *L. culinaris.* spp. *orientalis* provides valuable sources of drought tolerance gene for low rainfall environments like Syria, Jordan, Azerbaijan, Turkmenistan, and Tajikistan. Moreover, it is reported that interspecific offsprings are capable of drought tolerance that is associated with relative leaf water content, pubescent leaves, cell membrane stability, higher root-shoot ratio, higher transpiration, decreased wilting, and reduced canopy temperature (Omar et al. 2019). Recombinant inbred lines (RILs) developed from crosses of *L. culinaris* with *L. odemensis* and *L. orientalis* were evaluated by Sanderson et al. (2019). They found that tolerance to stress is associated with delayed flowering, reduced transpiration, and profound root system in the studied *Lens* species. Accessions of *L. orientalis and L. odemensis* with extensive roots showed tolerance to drought, whereas late flowering caused root extension deeper into the soil. The previous study emphasized that *L. tomentosus* showed a reduced transpiration rate, and *L. culinaris ssp. orientalis* and *L. lamottei* accession showed cold tolerance capacities that may be good candidates for lentil improvement breeding programs (Hamdi and Erskine 1996; Gorim and Vandenberg 2017).

Kumar et al. (2011, 2013) and Singh et al. (2013) identified variation of yield attributes and suitable donor for crop duration, biological yield, secondary branches, seed size, pod numbers, and yield in various *Lens* species. The accession ILWL 118 from *L. culinaris* ssp. *orientalis* is an excellent source of earliness that is critical for rice fallows and rainfed condition in Eastern and Central India, respectively. For

improvement in growth habit, biomass production, and seed traits, utilization of *L. ervoidis* as a donor source is reported (Tullu et al. 2013). *L. culinaris* ssp. *orientalis* and *L. lamottei* are the potential sources of genes pertaining to size of seed and seed and pod number (Ferguson et al. 1998; Gupta and Sharma 2006). Singh et al. (2014, 2020) evaluated and found significant variation for agronomical important traits like yield variation and resistance against multiple diseases in *L. ervoidis* and *L. nigricans* of *Lens* wild species. *L. culinaris* accessions JL 1, IPL 98/193, and DPL 53 showed excellent root parameters; therefore, these genotypes survive and sustain in drought conditions (Kumar et al. 2012). Malhotra et al. (2004) has reported that some lentil genotypes have strong mechanism to drought escape, namely ILL7618, ILL7981, ILL9922, ILL9830, ILL9844, ILL9920, ILL6024, ILL7504, ILL8095, ILL8138, and ILL8621 when research was performed in the ICARDA (Mediterranean environment). *Lens* crop wild relatives have developed rich genetic diversities and adapted a broad range of environments for drought tolerances (Hamdi and Erskine 1996). The primary approaches for combating drought stress are avoidance, escape, and tolerance in the evaluated *Lens* species (*L. culinaris* Eston, *L. odemensis* acc. IG 72623, *L. lamottei* acc. IG 110813, and L. *orientalis* acc. PI 572376, PI 572376) that are capable of tolerance to drought stress (Fang and Xiong 2015; Gorim and Vandenberg 2017). Recently studied by Rajendran et al. (2021), they marked the accession ILL 7835 as a considerably good source for stable tolerance against the combined stress of heat and drought under various environmental conditions in Morocco.

Major biotic stress donors for *Lens* species, i.e., Fusarium wilt, Stemphylium blight, Ascochyta blight, anthracnose, rust, powdery mildew, Sitona weevil, bruchids, and Orobanche, have been identified and utilized for improving cultivated lentil spp. (Meena et al. 2017). Moreover, 248 accessions of CWRs germplasm from ICARDA were evaluated for Ascochyta blight resistance by Bayaa et al. (1994). They found suitable donor in 3 accessions (*L. culinaris* ssp. *odemensis),* 12 accessions (*L. culinaris* ssp. *orientalis)*, and 36 accessions (*L. nigricans*) for Ascochyta blight resistance. Fernández-Aparicio et al. (2009) evaluated wild *Lens* accessions and identified resistance to broom rape in *Lens ervoides, Lens odemensis,* and *Lens orientalis*.

8 Brief Account of Molecular Mapping of HR Genes and QTLs

A few molecular mapping studies have been reported in lentil for nutritional traits (Table 6). In this section, we will be discussing them one by one. In a recent study, a set of 96 diverse lentil germplasm lines were evaluated at 3 different locations in India for iron and zinc concentrations, and the entire panel was genotypes with SSR markers (Singh et al. 2017). Association mapping found three simple sequence repeats (SSRs) (PBALC 13, PBALC 206, and GLLC 563) linked with grain Fe concentration (9–11% of phenotypic variation) and four SSRs (PBALC 353, SSR 317–1, PLC 62, and PBALC 217) were associated with grain Zn concentration (14–21% of phenotypic variation). These identified SSRs were found to be

Table 6 Molecular mapping of nutritional traits in lentil

Trait	QTLs	Phenotypic variance (%)	References
Iron concentration	3 MTAs	9–11%	Singh et al. 2017
	3 MTAs	6–13%	Kumar et al. 2019
	2 MTAs	9–21%	Khazaei et al. 2017
Zinc concentration	2 MTAs	6–17%	Kumar et al. 2019
	4 MTAs	14–21%	Singh et al. 2017
Selenium concentration	SeQTL2.1, SeQTL5.2, SeQTL5.1, SeQTL5.3	6–17%	Ates et al. 2016
Manganese concentration	MnQTL1.1, MnQTL1.2, MnQTL3.1, MnQTL3.2, MnQTL3.3, MnQTL7.1	15–24%	Ates et al. 2018

stable across locations. These candidate SSRs can be used in marker-assisted lentil breeding for iron and zinc concentration. In another study utilizing a similar approach using the linkage disequilibrium (LD) analysis with a mixed linear model (MLM), two SSR markers, GLLC 106 and GLLC 108, were associated with grain Fe concentration, explaining 17% and 6% phenotypic variation, respectively, and three SSR markers (PBALC 364, PBALC 92, and GLLC592) were associated with grain Zn concentration, explaining 6%, 8%, and 13% phenotypic variation, respectively (Kumar et al. 2019). Khazaei et al. (2017) used SNP genotyping of a set of 138 cultivated lentil accessions from 34 countries. The entire set was also phenotyped for iron and zinc concentration over four environments. The marker–trait association analysis detected two SNP markers tightly linked to seed Fe and one linked to seed Zn concentration($-\log10\, P \geq 4.36$). A few putative candidate genes were also detected for iron and zinc concentration. Ates et al. (2016) studied a panel of 96 recombinant inbred lines (RILs) developed from the cross "PI 320937" × "Eston" grown in three environments for 2 years. Se concentrations in seed varied between 119 and 883 μg/kg. Genotyping with 4 SSRs and 1780 SNPs and further statistical analysis marked 4 QTL regions and 36 putative QTL markers with seed Se concentration, explaining 6–17% of the total phenotypic variation. In another similar study, Ates et al. (2018) phenotyped a RIL population (120) (CDC Redberry × ILL7502) for manganese concentration. Genotyping with 5385 markers and further linkage analysis found a total of 6 QTL for Mn concentration that were identified using composite interval mapping (CIM). All QTL were statistically significant and explained 15–24% of the total phenotypic variation. Our group at IIPR, India, also identified molecular markers linked with genomic regions in lentil controlling high iron and zinc concentrations; very soon, this information will be in the public domain. It is imperative to mention that in most of the cases, the percent variation explained by linked markers at 0.05 or 0.01 level of significance is lower than their utilization in marker-assisted breeding. Hence, significantly linked SNPs

can be converted to PCR-based KASP (Kompetitive allele specific PCR) markers for routine use in these traits. Khazaei et al. (2017) reported that Fe and Zn concentrations were positively correlated; however, no common molecular marker could be found; hence, implementation of any marker-assisted selection would require independent selections.

9 Genetic Engineering for HR Traits

9.1 Traits of Interest and Foreign Genes

Genetic engineering methods are available today to incorporate alien genes which are usually not possible to be introgressed into cultivars due to hybridization barriers or incompatibility. With the interventions of molecular biology or genetic engineering today, alien gene coding for foreign proteins that improves the nutritional quality of a particular crop cultivar can be introduced, or already existing genes' expression may be enhanced or reduced, as in the case of anti-nutritional factors like phytic acid.

9.1.1 Iron Biofortification

Multiple methods to develop transgenic plants have been studied to enhance the iron concentration in different model species and cereal crops. However, we will be restricting our discussion on food legumes or pulses as far as possible.

Iron-Binding Protein Gene (Ferritin Gene)

Pea ferritin was found to be degraded in the case of exposure to gastric pH treatment, and the released iron was transported into the Caco-2 cells by DMT-1 (divalent metal transporter-1). It was further observed that inhibitors of DMT-1 and nonheme iron absorption reduced iron uptake by 26–40%. On the contrary, in the absence of gastric pH treatment, the iron uptake by pea ferritin was normal and unaffected by the inhibitors. Chlorpromazine (clathrin-mediated endocytosis inhibitor) has a negative impact on pea ferritin content; in the case of exposure to chlorpromazine, iron uptake can be reduced by ~30%, which further indicates that Fe is transported into cells via a clathrin-mediated endocytic pathway. In addition, 60% less ROS production could be found in pea ferritin as compared to $FeSO_4$. Few workers also reported that endosperm-specific expression of wheat and soybean ferritin in wheat led to several fold (1.5-fold for pea and 1.9-fold for soybean) increases in iron content, respectively (Borg et al. 2012; Xiaoyan et al. 2012). A few years back, in lentil, Sen Gupta et al. (2017) developed molecular markers for Ferritin-1, BHLH-1 (basic helix loop helix), or FER-like transcription factor protein and IRT-1 (iron-related transporter) genes utilizing genome synteny with barrel medic (*Medicago truncatula*). Significant expression of Ferritin-1 and IRT-1 was observed under excess iron conditions. More efforts are required for full-length cloning and functional validation of lentil ferritin gene for use in genetic engineering-mediated biofortification experiments in lentil.

Iron-Chelator Gene

Nicotianamine (NA) is involved in chelation of iron and regulate Fe homeostasis within the plant system (Curie et al. 2009). NA is the founding molecule of phytosiderophores, which are produced exclusively by the grass family, helping in the transport of Fe within plant system (Curie et al. 2009).

Iron Reductase Gene

Iron reductase genes reduce ferric ions to ferrous form. This reduction step is required for phytosiderophores-mediated uptake of iron (Schröder et al. 2003). Overexpressing iron reductase gene can help in increased uptake in iron from soil (Douchkov et al. 2005). Pea (brz and dgl) and Arabidopsis (frd3/man1) mutants expressing iron reductase accumulate higher concentration of Fe. In fact, among these mutants, high nicotianamine levels were also observed (Rogers and Guerinot 2002).

Insertion of Transporter Gene

Iron is transported via different plant transporters through root-soil interface and get stored in apoplasm (Morrissey and Guerinot 2009). The metal transporters of lentil may be studied on this aspect based on the clues established in model species and cereals (Kim et al. 2006; Connorton et al. 2017). Sen Gupta et al. (2017) developed molecular marker specific to IRT-1 (iron-related transporter) transporter gene and also observed genotypic variability in gene expression analysis for this gene; full-length cloning and functional analysis are required for this and other heavy metal transporter gene families.

Decreasing Anti-nutrient

Phytic acid is known to be an anti-nutrient which impairs the iron and zinc absorption or bioavailability after consumption. Transgenic wheat plants show an almost fourfold increase in phytase activity (Brinch-Pedersen et al. 2000). In another study, transgenic soybean expressed 2.5-fold higher activity of phytase (Gao et al. 2007). Synthetic phytase gene construct accumulated a higher quantity of proteins than the native ones (Kohli et al. 2006).

Increasing Enhancers for Increased Fe Absorption

Some dietary components have been known to increase iron absorption. These include vitamins such as β-carotene, ascorbic acid, α-tocopherol, and amino acids, which are released from proteins during digestion. Ascorbic acid and citric acid are known to reduce Fe to a ferrous state and improve absorption in the small intestine. Therefore, transgenic approaches can be used to overexpress ascorbic acid in combination with ferritin (Gropper et al. 2006). An increased cysteine content has also been shown to have a good effect on Fe absorption (Layrisse et al. 1984).

9.1.2 Zinc Biofortification

Zinc uptake, transport, and accumulation are similar to iron metabolism; however, more efforts are required to understand the zinc-specific transporter gene families (Zimmermann and Hurrell 2002). It is prevalent worldwide due to widespread soil

deficiency. Hence, there is a strong demand to increase zinc concentration in staple crops. Improving bioavailability is important in this case as this can make use of even lower uptake (Lonnerdal 2003).

Overexpression of NAS Gene Family

Nicotianamine (NA) is a ubiquitous chelator of transition metals which transports iron and zinc within plant system. NA synthase (NAS) enzyme is involved in the synthesis of NA from S-adenosylmethionine (Takahashi et al. 2003). Increasing NA concentration in a plant through transgenics by the overexpression of NAS genes showed optimistic results in many cereals.

Overexpression of NAC Gene Family

NAC transcription factors have a critical role in mobilizing iron and zinc during senescence. By increasing the transcription factors activity, the iron and zinc mobilization may be improved, as was observed in transgenics that showed increased accumulation of zinc (Connorton et al. 2017). In the dicot plant lentil, the importance of these transcription factors has to be studied in detail.

9.2 Achievements of Transgenics in Lentil

Transgenics or genetic engineering technologies are providing opportunity to improve agronomical traits as well as micronutrient concentration in legumes. *Agrobacterium*-mediated genetic transformation has been considered the most common and successful method for trait improvement in grain legumes like pea and soybean (Schroeder et al. 1993). Protoplasts cannot be transformed with *Agrobacterium*, but they can be cloned, whereas single-event transformants takes a long time to regenerate into plants. However, gene gun (particle bombardment) is efficient method to any plant tissue, but has the unpredictability of gene integration and high risks. Also, *Agrobacterium*-mediated transformation in many legumes are difficult due to their susceptibility to *Agrobacterium* infection and very low transformation efficiency (0.03–5.1%) (Yan et al. 2000). The efficiency of explant tissues (cotyledonary, embryo with single cotyledonary disc and node, and decapitated embryo) towards transformation through GUS (β-glucuronidase) found to express the GUS gene following histochemical assay. The explants showed inconsistent nature of GUS expression. In the case of lentil, explants showed much greater areas with GUS expression while some studied samples showed a small portion of the wounded cells competent for transformation (Warkentin and McHughen 1992). However, genetic transformation (*Agrobacterium*-mediated) is also influenced by several factors (Hashem 2007), including bacterial strain, duration of cocultivation time, explant type, etc.

Warkentin and McHughen (1992) reported that inoculation of lentil epicotyl explants for 10–15 min found to be suitable for GUS-positive putative explants. Whereas, many researchers reported that longer coculture period is capable of enhancing the GUS-infected area in lentil explants (Hashem 2007). The virulence

of the bacterial strain on lentil shoot apices explants differs significantly (Warkentin and McHughen 1992).

9.3 Prospects of Cisgenics

Biotechnological approaches based plant improvement is a continuous process that alters heritability of the trait and induces variation. Therefore, cisgenesis and intragenesis are utilized to create genetic variation in cultivated or existing germplasm for improving their quality and quantity. The advancement of sequencing technologies and genome information facilitates the isolation of intact Cis-genes with associated promoter/terminator from wild or cultivated species, which are utilized to insert into the genome of closely related and crossable species. In the case of intragenesis, different coding and regulatory sequences are assembled either in sense or in antisense orientation.

The applications and future prospects of cisgenesis and intragenesis in the improvement of many crops have been studied (Singh et al. 2018b). Disease-resistant cultivar like late blight-resistant potato (gene Rpi-sto1, Rpi-vnt1.1) and scab-resistant apple has been developed. Thus, cisgenesis as a powerful approach to transfer gene of interest without linkage drag. Cisgenic- and intragenic-derived genetically modified plants (GMP) are eco-friendly as classical breeding plants and also exempted from GMP legislation. Dudziak et al. (2019) suggested that application of novel scientific approaches is of major importance for improving the crop plants. However, these most efficient strategies based on genetic modification are still very controversial issues. Opponents of GMO crops do not agree with the use of genetic engineering in crop improvement and the production of new varieties suited for organic agriculture.

Cisgenesis suffers many a times from the random insertion in the genome causing variation in gene expression (Cardi 2016). Also, the random insertion may cause silencing of other genes. Random cisgene integration is similar to transgenic varieties, natural transposons, and induced translocations (Schmidt 2002). Another issue is the copy number and presence of vector sequence while transferred into a recipient genome. Schouten et al. (2006) reported that cisgenic transformation through *Agrobacterium* may also transfer small T-DNA borders. Approximately 80% of plants regenerated from cisgenic transformation experiments with vector backbone sequences. Projects for increasing iron and zinc in crops such as lentil are at varied stages of development (Saltzman et al. 2013). Cisgenesis has a long way to go in lentil improvement since, on the issue of safety, regulators could treat cisgenic plants the same as conventionally bred plants (Schouten et al. 2006).

10 Future Prospects and Conclusion

Lentil is a highly nutritional crop species having a tremendous opportunity as a biofortification crop. Already worldwide biofortified lentil varieties are being grown by the growers. However, there is still scope to better understand metabolism of

biofortification-related traits in lentil. Only two traits, iron and zinc concentration, have been addressed so far; other important traits, such as folate concentration, Se concentration, and other micronutrients, can be investigated. Profiling of wild or related species as new sources of biofortification related traits is required. Moreover, more studies are required to map the biofortification traits in the lentil genome for their use in breeding programs. New tools like haplotype breeding using next-generation sequencing platforms have a great potential for use in this crop species. As far as transgenic or cisgenic research is concerned, functional analysis of genes related to biofortification related traits is a priority area. Lastly, the human nutrition component needs to be included in lentil biofortification research to address the malnutrition in a more holistic way.

References

Amarakoon D, Thavarajah D, McPhee K, Thavarajah P (2012) Iron-, zinc-, and magnesium-rich field peas (*Pisum sativum* L.) with naturally low phytic acid: a potential food-based solution to global micronutrient malnutrition. J Food Compos Anal 27:8–13

Amiri R, Bahraminejad S, Cheghamirza K, Arzani A (2020) Genetic analysis of iron and zinc concentrations in bread wheat grains. Journal of Cereal Science 95:103077

Ates D, Sever T, Aldemir S, Yagmur B, Temel HY, Kaya HB et al (2016) Identification QTLs controlling genes for Se uptake in lentil seeds. PLoS One 15:e0149210. https://doi.org/10.1371/journal.pone.0149210

Ates D, Aldemir S, Yagmur B, Kahraman A, Ozkan H, Vandenberg A et al (2018) QTL mapping of genome regions controlling manganese uptake in lentil seed. G3: Genes Genomes Genet 8: 1409–1416. https://doi.org/10.1534/g3.118.200259

Batra J, Seth PK (2002) Effect of iron deficiency on developing rat brain. Indian J Clin Biochem 17(2):108–114

Baum M, Hamweih A, Furman B, Sarker A, Erskine W (2008) Biodiversity, genetic enhancement and molecular approaches in lentils. In: Sharma AK, Sharma A (eds) Plant genome, biodiversity and evolution. Science Publishers, Enfield, pp 135–152

Bayaa B, Erskine W, Hamdi A (1994) Response of wild lentil to Ascochyta fabae f. sp. lentis from Syria. Genetic Resources and Crop Evolution 41:61–5

Biswas S, Talukder G, Sharma A (1999) Prevention of cytotoxic effects of arsenic by short-term dietary supplementation with selenium in mice in vivo. Mutat Res/Genet Toxicol Environ Mutagen 441(1):155–160

Blair MW (2013) Mineral biofortification strategies for food staples: the example of common bean. Agric Food Chem 61:8287–8294

Borg S, Brinch-Pedersen H, Tauris B, Madsen LH, Darbani B, Noeparvar S et al (2012) Wheat ferritins: improving the iron content of the wheat grain. J Cereal Sci 56:204–213. https://doi.org/10.1016/j.jcs.2012.03.005

Borrill P, Connorton JM, Balk J, Miller AJ, Sanders D, Uauy C (2014) Biofortification of wheat grain with iron and zinc: integrating novel genomic resources and knowledge from model crops. Front Plant Sci 5:53. https://doi.org/10.3389/fpls.2014.00053

Bouis HE, Hotz C, McClafferty B, Meenakshi JV, Pfeiffer WH (2011) Biofortification: a new tool to reduce micronutrient malnutrition. Food Nutr Bull 32:31S–40S

Brinch-Pedersen H, Olesen A, Rasmussen SK, Holm PB (2000) Generation of transgenic wheat (Triticum aestivum L.) for constitutive accumulation of an Aspergillus phytase. Molecular Breeding 6:195–206

Cakmak I (2008) Enrichment of cereal grains with zinc: agronomic or genetic biofortification? Plant Soil 302:1e17

Cakmak I, Ozkan H, Braun H-J, Welch RM, Romheld V (2000) Zinc and iron concentrations in seeds of wild primitive and modern wheats. Food Nutr Bull 21:e401–e403
Cardi T (2016) Cisgenesis and genome editing: combining concepts and efforts for a smarter use of genetic resources in crop breeding. Plant Breed 135(2):139–147. https://doi.org/10.1111/pbr.12345
Connorton JM, Balk J, Rodríguez-Celma J (2017) Iron homeostasis in plants–a brief overview. Metallomics 9(7):813–23
Cubero JI, Perez De La Vega M, Fratini R (2009) Origin, phylogeny, domestication and spread. InThe lentil: botany, production and uses (pp. 13–33). Wallingford UK: CABI
Curie C, Gaëlle C, Daniel C, Fanchon D, Kyoko H, Marie Le J, Julie M, Adam S, Pierre C, Stéphane M (2009) Metal movement within the plant: contribution of nicotianamine and yellow stripe 1-like transporters. Ann Bot 103:1–11
Darai R, Sarker A, Kumar S, Dhakal K, Neupane R, Aryal L, Poudel P (2020) Lentil improvement in Nepal
DellaValle DM, Thavarajah D, Thavarajah P, Vandenberg A, Glahn RP (2013) Lentil (*Lens culinaris* L.) as a candidate crop for iron biofortification: is there genetic potential for iron bioavailability. Field Crops Res 144:119–125
Dikshit HK, Mishra GP, Aski MS, Singh A, Tripathi K, Bansal R, Pratap A, Gupta S, Kumar S (2022) Lentil breeding. In Fundamentals of field crop breeding (pp. 1181–1236). Singapore: Springer Nature Singapore
Douchkov D, Gryczka C, Stephan UW, Hell R, Bäumlein H (2005) Ectopic expression of nicotianamine synthase genes results in improved iron accumulation and increased nickel tolerance in transgenic tobacco. Plant, Cell & Environment 28(3):365–74
Doyle JJ (1988) 5S ribosomal gene variation in the soybean and its progenitor. Theor Appl Genet 75:621–624
Dudziak K, Magdalena S, Krzysztof K, Michał N (2019) Cisgenesis as a novel prospect for crop improvement. A review. Agron Sci. https://doi.org/10.24326/as.2019.2.1
Fang Y, Xiong L (2015) General mechanisms of drought response and their application in drought resistance improvement in plants. Cellular and Molecular Life Sciences 72:673–89
FAOSTAT (2007). http://faostat.fao.org/
FAOSTAT (2011). http://faostat.fao.org/
Ferguson ME, Robertson LD, Ford-Lloyd BV, Newbury HJ, Maxted N (1998) Contrasting genetic variation amongst lentil landraces from different geographical origins. Euphytica 102:265–73
Fernández-Aparicio M, Flores F, Rubiales D (2009) Recognition of root exudates by seeds of broomrape (Orobanche and Phelipanche) species. Annals of Botany 103(3):423–31
Fratini R, Ruiz ML (2006) Interspecific hybridization in the genus Lens applying in vitro embryo rescue. Euphytica 150:271–280
Fratini R, Ruiz ML, Perez de la Vega M (2004) Intra-specific and inter-sub-specific crossing in lentil (*Lens culinaris* Medik). Can J Plant Sci 84:981–986
Frossard E, Bucher M, Mächler F, Mozafar A, Hurrell R (2000) Potential for increasing the content and bioavailability of Fe, Zn and Ca in plants for human nutrition. J Sci Food Agric 80(7):861–879
Gao XR, Wang GK, Su Q, Wang Y, An LJ (2007) Phytase expression in transgenic soybeans: stable transformation with a vector-less construct. Biotechnology Letters 29:1781–7
Gomez-Galera S, Rojas E, Sudhakar D, Zhu CF, Pelacho AM, Capell T et al (2010) Critical evaluation of strategies for mineral fortification of staple food crops. Transgenic Res 19:165–180. https://doi.org/10.1007/s11248-009-9311-y
Gorim LY, Vandenberg A (2017) Evaluation of wild lentil species as genetic resources to improve drought tolerance in cultivated lentil. Frontiers in plant science. 8:1129
Graham RD, Rengel Z (1993) Genotypic variation in zinc uptake and utilization. In: Robson AD (ed) Zinc in soils and plants. Kluwer Academic, Dordrecht, pp 107–118
Graham RD, Welch RM, Saunders DA, Ortiz-Monasterio I, Bouis HE, Bonierbale M, de Haan S, Burgos G, Thiele G, Liria R, Meisner CA, Beebe SE, Potts MJ, Kadian M, Hobbs PR, Gupta RK, Twomlow S (2007) Nutritious subsistence food systems. Adv Agron 92:1–74

Gropper SS, Blessing D, Dunham K, Barksdale JM (2006) Iron status of female collegiate athletes involved in different sports. Biological Trace Element Research 109:1–3

Guerra-García A, Gioia T, von Wettberg E, Logozzo G, Papa R, Bitocchi E, Bett KE (2021) Intelligent characterization of lentil genetic resources: Evolutionary history, genetic diversity of germplasm, and the need for well-represented collections. Curr Protoc 1:e134. https://doi.org/10.1002/cpz1.134

Gupta D, Sharma SK (2006) Evaluation of wild Lens taxa for agro-morphological traits, fungal diseases and moisture stress in North Western Indian Hills. Genetic Resources and Crop Evolution 53:1233–41

Gupta DS, Thavarajah D, Knutson P, Thavarajah P, McGee RJ, Coyne CJ, Kumar S (2013) Lentils (*Lens culinaris* L) a rich source of folates. J Agric Food Chem 61:7794–7799

HarvestPlus (2014) Biofortification progress briefs: iron and zinc lentils. Brief 9, pp 19. Available online at www.HarvestPlus.org

Hashem R (2007) Improvement of lentil (*Lens culinaris* Medik.) through genetic transformation. PhD thesis, Dhaka

Hawkesford MJ, Zhao FJ (2007) Strategies for increasing the selenium content of wheat. J Cereal Sci 46(3):282–292

Holmberg RE Jr, Ferm VH (1969) Interrelationships of selenium, cadmium, and arsenic in mammalian teratogenesis. Arch Environ Health 18(6):873–877

Hotz C, Brown KH, Rivera JA et al (2004) Assessment of the risk of zinc deficiency in populations and options for its control. Food Nutr Bull 25(Suppl 2):S94–S204

Johnson C, Thavarajah D, Thavarajah P (2013a) The influence of phenolic and phytic acid food matrix factors on iron bioavailability potential in 10 commercial lentil genotypes (*Lens culinaris* L). J Food Compos Anal 31(1):82–86. https://doi.org/10.1016/j.jfca.2013.04.003

Johnson CR, Thavarajah D, Combs GF Jr, Thavarajah P (2013b) Lentil (*Lens culinaris* L.): a prebiotic-rich whole food legume. Food Res Int 51:107–113

Joshi AK, Crossa I, Arun B, Chand R, Trethowan R, Vargas M, Ortiz-Monasterio I (2010) Genotype _ environment interaction for zinc and iron concentration of wheat grain in eastern Gangetic plains of India. Field Crops Res 116:268–277

Karakoy T, Halil E, Baloch FS, Toklu F, Eker S, Kilian B, Ozkan H (2012) Diversity of macro-andmicronutrients in the seeds of lentil landraces. Sci World J 2012(2):710412. https://doi.org/10.1100/2012/710412

Khan MA, Fuller MP, Baloch FS (2008) Effect of soil applied zinc sulphate on wheat (*Triticum aestivum* L.) grown on a calcareous soil in Pakistan. Cereal Res Commun 36(4):571–582

Khazaei H, Podder R, Caron CT, Kundu SS, Diapari M, Vandenberg A et al (2017) Marker–trait association analysis of iron and zinc concentration in lentil (Medik.) seeds. Plant Genome 10. https://doi.org/10.3835/plantgenome2017.02.0007

Kim SA, Punshon T, Lanzirotti A, Li L, Alonso JM, Ecker JR, Kaplan J, Guerinot ML (2006) Localization of iron in Arabidopsis seed requires the vacuolar membrane transporter VIT1. Science 314(5803):1295–8

Kohli A, González-Melendi P, Abranches R, Capell T, Stoger E, Christou P (2006) The quest to understand the basis and mechanisms that control expression of introduced transgenes in crop plants. Plant Signaling & Behavior 1(4):185–95

Kumar SH, Kumar JI, Singh RG, Sarke A, Baum MI (2011) Lentils for dryland agriculture: present status and future strategy. Flora and Fauna 17(1):100–7

Kumar J, Basu PS, Srivastava E, Chaturvedi SK, Nadarajan N, Kumar S (2012) Phenotyping of traits imparting drought tolerance in lentil. Crop and Pasture Science 63(6):547–54

Kumar S, Barpete S, Kumar J, Gupta P, Sarker A (2013) Global lentil production: constraints and strategies. SATSA Mukhapatra-Annual Technical 17:1–3

Kumar H, Singh A, Jain N, Kumari J, Singh AM, Singh D, Sarker A, Prabhu KV (2014a) Characterization of grain iron and zinc in lentil (*Lens culinaris* Medikus culinaris) and analysis of their genetic diversity using SSR markers. Aust J Crop Sci 8:1005–1012

Kumar J, Kant R, Quatadah SM, Kumar S, Sarker A (2016a) Biofortification through breeding interventions in lentil. In: Singh U, Praharaj C, Singh S, Singh N (eds) Biofortification of Food Crops. Springer, New Delhi. https://doi.org/10.1007/978-81-322-2716-8_12

Kumar J, Gupta DS, Kumar S, Gupta S, Singh NP (2016b) Current knowledge on genetic biofortification in lentil. Journal of Agricultural and Food Chemistry 64(33):6383–96

Kumar J, Thavarajah D, Kant R, Sarker A, Singh NP (2016c) Genetic variability for mineral nutrients in lentil. J Food Legumes 29(2):94–101

Kumar J, Kumar S, Sarker A, Singh NP (2018a) Analysis of genetic variability and genotype× environment interactions for iron and zinc content among diverse genotypes of lentil. J Food Sci Technol 55(9):3592–3605

Kumar S, Choudhary AK, Rana KS, Sarker A, Singh M (2018b) Bio-fortification potential of global wild annual lentil core collection. PLoS One 13(1):e0191122

Kumar H, Singh A, Dikshit HK, Mishra GP, Aski M, Meena MC et al (2019) Genetic dissection of grain iron and zinc concentrations in lentil (*Lens culinaris* Medik.). J Genet 98:1–4. https://doi.org/10.1007/s12041-019-1112-3

Ladizinsky G (1979) The origin of lentil and its wild genepool. Euphytica 28:179–187

Layrisse M, Martínez-Torres C, Leets I, Taylor P, Ramírez J (1984) Effect of histidine, cysteine, glutathione or beef on iron absorption in humans. The Journal of Nutrition 114(1):217–23

Levander OA (1977) Metabolic interrelationships between arsenic and selenium. Environ Health Perspect 19:159

Lonnerdal B (2003) Genetically modified plants for improved trace element nutrition. The Journal of Nutrition 133(5):1490S–3S

Lyons G, Ortiz-Monasterio I, Stangoulis J, Graham R (2005) Selenium concentration in wheat grain: is there sufficient genotypic variation to use in breeding? Plant Soil 269:369–380

Meena JK, Singh A, Dikshit HK, Mishra GP, Aski M, Srinivasa N, Gupta S, Singh D, Tripathi A (2017) Screening of Lentil (Lens culinaris Medikus sub sp. culinaris) Germplasm against Fusarium Wilt (*Fusarium oxysporum* f. sp. lentis). International Journal Current Microbiol Applied Science 6(11):2533–41

Miller DD, Welch RM (2013) Food system strategies for preventing micronutrient malnutrition. Food Policy 42:115–28

Morrissey J, Guerinot ML (2009) Iron uptake and transport in plants: the good, the bad, and the ionome. Chemical Reviews 109(10):4553–67

Muehlbauer FJ, Cho S, Sarker A, McPhee KE, Coyne CJ, Rajesh PN, Ford R (2006) Application of biotechnology in breeding lentil for resistance to biotic and abiotic stress. Euphytica 147: 149–165

Nestel P, Bouis HE, Meenakshi JV, Pfeiffer W (2006) Biofortification of staple food crops. J Nutr 136(4):1064–1067

Ogutcen E, Ramsay L, von Wettberg EB, Bett KE (2018) Capturing variation in Lens (Fabaceae): Development and utility of an exome capture array for lentil. Applications in Plant Sciences 6(7):e01165

Omar O, Elgali I, Dahlin C, Thomsen P (2019) Barrier membranes: More than the barrier effect?. Journal of Clinical Periodontology 46:103–23

Ortiz-Monasterio I, Palacios-Rojas N, Meng E, Pixley K, Trethowan R, Pena RJ (2007) Enhancing the mineral and vitamin content of wheat and maize through plant breeding. J Cereal Sci 46: 293e307

Petry N, Boy E, Wirth JP, Hurrell RF (2015) Review: the potential of the common bean (*Phaseolus vulgaris*) as a vehicle for iron biofortification. Nutrients 7:1144–1173

Prentice AM, Gershwin ME, Schaible UE, Keusch GT, Victoria LG, Gordon JI (2008) New challenges in studying nutrition disease interactions in the developing world. J Clin Invest 118:1322e1329

Rahman M, Erskine W, Zaman MS, Thavarajah P, Thavarajah D, Siddique K (2013) Selenium biofortification in lentil (*Lens culinaris* Medikus subsp. culinaris): farmers' field survey and

genotype × environment effect. Food Res Int 54:1596–1604. https://doi.org/10.1016/j.foodres.2013.09.008

Rajendran K, Coyne CJ, Zheng P, Saha G, Main D, Amin N, Ma Y, Kisha T, Bett KE, McGee RJ, Kumar S (2021) Genetic diversity and GWAS of agronomic traits using an ICARDA lentil (Lens culinaris Medik.) Reference Plus collection. Plant Genetic Resources 19(4):279–88

Reilly C (1996) Selenium in food and health. Blackie, London

Rogers EE, Guerinot ML (2002) FRD3, a member of the multidrug and toxin efflux family, controls iron deficiency responses in Arabidopsis. The Plant Cell 14(8):1787–99

Saltzman A, Birol E, Bouis HE, Boy E, DeMoura FF, Islam Y, Pfeiffer WH (2013) Biofortification: progress toward a more nourishing future. Glob Food Secur 2:9–17

Sanderson LA, Caron CT, Tan R, Chan C, Bett K (2019) Knowpulse: an evolving breeder-friendly web-portal for chickpea, common bean, field pea and lentil. InPlant and Animal Genome XXVII Conference (January 12–16, 2019) 2019 Jan 12. PAG

Sarker A, El-Askhar F, Uddin MJ, Million E, Yadav NK, Dahan R, Wolfgang P (2007) Lentil improvement for nutritional security in the developing world. Paper presented at the ASA-CSSA-SSSA international annual meetings, New Orleans, LA, 4–8 Nov 2007

Schmidt R (2002) Plant genome evolution: lessons from comparative genomics at the DNA level. Plant Mol Biol 48:21–37. https://doi.org/10.1007/978-94-010-0448-0_2

Schouten HJ, Krens FA, Jacobsen E (2006) Cisgenic plants are similar to traditionally bred plants: international regulations for genetically modified organisms should be altered to exempt cisgenesis. EMBO Rep 7(8):750–753. https://doi.org/10.1038/sj.embor.7400769

Schreier AD, Mahardja B, May B (2012) Hierarchical patterns of population structure in the endangered Fraser River white sturgeon (*Acipenser transmontanus*) and implications for conservation. Can J Fish Aquat Sci 69:1968–1980

Schröder I, Johnson E, De Vries S (2003) Microbial ferric iron reductases. FEMS Microbiology Reviews 27(2–3):427–47

Schroeder HE, Schotz AH, Wardley-Richardson T, Spencer D, Higgins TJ (1993) Transformation and regeneration of two cultivars of pea (Pisum sativum L.). Plant Physiology 101(3):751–7

Sen Gupta D, Thavarajah D, McGee RJ, Coyne CJ, Kumar S, Thavarajah P (2016) Genetic diversity among cultivated and wild lentils for iron, zinc, copper, calcium and magnesium concentrations. Australian Journal of Crop Science 10(10):1381–7

Sen Gupta D, McPhee K, Kumar S (2017) Development of molecular markers for iron metabolism related genes in lentil and their expression analysis under excess iron stress. Front Plant Sci 8:579

Singh M, Rana MK, Kumar K, Bisht IS, Dutta M, Gautam NK, Sarker A, Bansal KC (2013) Broadening the genetic base of lentil cultivars through inter-sub-specific and interspecific crosses of L ens taxa. Plant Breeding 132(6):667–75

Singh M, Bisht IS, Kumar S, Dutta M, Bansal KC, Karale M, Sarker A, Amri A, Kumar S, Datta SK (2014) Global wild annual Lens collection: a potential resource for lentil genetic base broadening and yield enhancement. PLoS One 9(9):e107781

Singh A, Sharma V, Dikshit HK, Aski M, Kumar H, Thirunavukkarasu N, Patil BS, Kumar S, Sarker A (2017) Association mapping unveils favorable alleles for grain iron and zinc concentrations in lentil (*Lens culinaris* subsp. *culinaris*). PLoS One 12(11):e0188296. https://doi.org/10.1371/journal.pone.0188296

Singh V, Singh S, Kumari S, Kumar A (2018b) Cisgenesis a sustainable approach of gene introgression and its utilization in horticultural crops: a review. Int J Curr Microbiol App Sci 7:5002–5009

Singh M, Kumar S, Basandrai AK, Basandrai D, Malhotra N, Saxena DR, Gupta D, Sarker A, Singh K (2020) Evaluation and identification of wild lentil accessions for enhancing genetic gains of cultivated varieties. PLoS One 15(3):e0229554

Smrkolj P, Germ M, Kreft I, Stibilj V (2006) Respiratory potential and Se compounds in pea (*Pisum sativum* L.) plants grown from Se-enriched seeds. J Exp Bot 57(14):3595–3600

Spallholz JE, Boylan LM, Rhaman MM (2004) Environmental hypothesis: is poor dietary selenium intake an underlying factor for arsenicosis and cancer in Bangladesh and West Bengal, India? Sci Total Environ 323(1–3):21–32

Spallholz JE, Mallory Boylan L, David Robertson J, Smith L, Rahman MM, Hook J, Rigdon R (2008) Selenium and arsenic content of agricultural soils from Bangladesh and Nepal. Toxicol Environ Chem 90(2):203–210

Stewart CP, Dewey KG, Ashoran P (2010) The undernutrition epidemic: an urgent health priority. Lancet 375:282

Tahir M, Vandenberg A, Chibbar RN (2011) Influence of environment on seed soluble carbohydrates in selected lentil cultivars. J Food Compos Anal 24(4–5):596–602

Takahashi M, Terada Y, Nakai I, Nakanishi H, Yoshimura E, Mori S, Nishizawa NK (2003) Role of nicotianamine in the intracellular delivery of metals and plant reproductive development. The Plant Cell 15(6):1263–80

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps unlocking genetic potential from the wild. Science 277:1063–1066

Terry N, Zayed AM, De Souza MP, Tarun AS (2000) Selenium in higher plants. Annu Rev Plant Biol 51(1):401–432

Thavarajah D, Vandenberg A, George GN, Pickering IJ (2007) Chemical form of selenium in naturally selenium-rich lentils (*Lens culinaris* L.) from Saskatchewan. J Agric Food Chem 55: 7337–7341. https://doi.org/10.1021/jf070681i

Thavarajah D, Ruszkoski J, Vandenberg A (2008) High potential for selenium bio-fortification of lentils (*Lens culinaris* L.). J Agric Food Chem 56:10747–10753

Thavarajah D, Thavarajah P, Sarker A, Vandenberg A (2009a) Lentils (*Lens culinaris* Medikus subsp. *culinaris*): a whole food for increased iron and zinc intake. J Agri Food Chem 57:5413–5419

Thavarajah P, Thavarajah D, Vandenberg A (2009b) Low phytic acid lentils (*Lens culinaris* L.): a potential solution for increased micronutrient bioavailability. J Agri Food Chem 57:9044–9049

Thavarajah D, Thavarajah P, Wejesuriya A, Rutzke M, Glahn RP, Combs GF Jr, Vandenberg A (2011a) The potential of lentil (*Lens culinaris* L) as a whole food for increased selenium, iron and zinc intake: preliminary results from a 3 year study. Euphytica 180:123–128

Thavarajah D, Thavarajah P, Sarker A, Michael M, George V, Renuka S, Omar I, Omar H, Beybin B, Albert V (2011b) A global survey of effects of genotype and environment on selenium concentration in lentils (*Lens culinaris* L.): Implications for nutritional fortification strategies. Food Chem 125(2011):72–76

Thavarajah D, Thavarajah P, Vial E, Gebhardt M, Lacher C, Kumar S, Combs GF (2015) Will selenium increase lentil (*Lens culinaris* Medik) yield and seed quality? Front Plant Sci 6:356. https://doi.org/10.3389/fpls.2015.00356

Tripathi K, Kumari J, Gore PG, Mishra DC, Singh AK, Mishra GP, Gayacharan C, Dikshit HK, Singh N, Semwal DP, Mehra R (2022) Agro-morphological characterization of lentil germplasm of Indian national genebank and development of a core set for efficient utilization in lentil improvement programs. Frontiers in Plant Science 12:751429

Tullu A, Bett K, Banniza S, Vail S, Vandenberg A (2013) Widening the genetic base of cultivated lentil through hybridization of Lens culinaris 'Eston' and L. ervoides accession IG 72815. Canadian Journal of Plant Science 93(6):1037–47

Turakainen M (2007) Selenium and its effects on growth, yield and tuber quality in potato. University of Helsinki, Department of Applied Biology, Publication 30. Helsinki, 50 p. + appendices

Velu G, Ortiz-Monasterio I, Cakmak I, Hao Y, Singh RP (2014) Biofortification strategies to increase grain zinc and iron concentrations in wheat. J Cereal Sci 59:365e372

Wang N, Hatcher DW, Toews R, Gawalko EJ (2009) Influence of cooking and dehulling on nutritional composition of several varieties of lentils (*Lens culinaris*). Food Sci Technol 42(4):842–848

Warkentin TD, McHughen A (1992) Agrobacterium tumefaciens-mediated beta-glucuronidase (GUS) gene expression in lentil (Lens culinaris Medik.) tissues. Plant Cell Reports 11:274–8

Welch MR, Graham DR (1999) A new paradigm for world agriculture: meeting human needs productive sustainable nutritious. Field Crops Res 60:1–10

Wessells KR, Brown KH (2012) Estimating the global prevalence of zinc deficiency: results based on zinc availability in national food supplies and the prevalence of stunting. PLoS One 7(11): e505–e568

White PJ, Broadley MR (2009) Biofortification of crops with seven mineral elements often lacking in human diets -iron, zinc, copper, calcium, magnesium, selenium and iodine. New Phytol 182(1):49–84

WHO (2008) The world health report. World Health Organization, Geneva

WHO (2009) The world health report. World Health Organization, Geneva

WHO (2012) The world health report. World Health Organization, Geneva

Wong MML, Gujaria-Verma N, Ramsay L, Yuan HY, Caron C, Diapari M et al (2015) Classification and characterization of species within the genus lens using genotyping-by-sequencing (GBS). PLoS One 10(3):e0122025. https://doi.org/10.1371/journal.pone.0122025

Wright DM, Neupane S, Heidecker T, Haile TA, Chan C, Coyne CJ, McGee RJ, Udupa S, Henkrar F, Barilli E, Rubiales D (2021) Understanding photothermal interactions will help expand production range and increase genetic diversity of lentil (Lens culinaris Medik.). Plants, People, Planet 3(2):171–81

Xiaoyan S, Yan Z, Shubin W (2012) Improvement Fe content of wheat (*Triticum aestivum*) grain by soybean ferritin expression cassette without vector backbone sequence. J Agric Food Biochem 20:766–773. Available online at: https://www.semanticscholar.org/paper/Improvement-Fe-Content-of-Wheat(Triticum-aestivum)-Xiao/74c0480abe93478866cc155188e64407f1ff68e6

Yan B, Srinivasa Reddy MS, Collins GB, Dinkins RD (2000) Agrobacterium tumefaciens–mediated transformation of soybean [Glycine max (L.) Merrill.] using immature zygotic cotyledon explants. Plant Cell Reports 19:1090–7

Yang XE, Chen WR, Feng Y (2007) Improving human micronutrient nutrition through biofortification in the soil plant system: China as a case study. Environ Geochem Health 29: 413–428

Zimmermann MB, Hurrell RF (2002) Improving iron, zinc and vitamin A nutrition through plant biotechnology. Current Opinion in Biotechnology 13(2):142–5

Grain Micronutrients in Pigeonpea: Genetic Improvement Using Modern Breeding Approaches

Aloleca Mukherjee, Anjan Hazra, Dwaipayan Sinha, Prathyusha Cheguri, Shruthi H B, Sanatan Ghosh, Naresh Bomma, Rituparna Kundu Chaudhuri, Prakash I. Gangashetty, and Dipankar Chakraborti

Contents

A. Mukherjee · A. Hazra · S. Ghosh · D. Chakraborti (✉)
Department of Genetics, University of Calcutta, Kolkata, West Bengal, India
e-mail: dcgntcs@caluniv.ac.in

D. Sinha
Department of Botany, Government General Degree College Mohanpur, Mohanpur, West Bengal, India

P. Cheguri · S. H. B · N. Bomma · P. I. Gangashetty (✉)
Pigeonpea Improvement Program, International Crops Research Institute for Semi-Arid Tropics (ICRISAT), Patancheru, Telangana, India
e-mail: p.gangashetty@cgiar.org

R. Kundu Chaudhuri
Department of Botany, Barasat Government College, Barasat, West Bengal, India

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_28

Abstract

The green revolution increased crop productivity and significantly reduced starvation and protein malnutrition. However, this caused micronutrient depleted soil, thereby responsible for widespread deficiencies of plant nutrients. Legumes are the important constituents of traditional healthy diets worldwide and second in agricultural importance after cereals. On a worldwide scale, pigeonpea ranks sixth among all legume crops and is India's second most important legume. Biofortification is the process of enhancing the nutrient value of crops using conventional selective breeding and agronomic approaches or via genetically modifying them. In many Indian states, the seeds of pigeonpea serve as a protein-rich pulse and are consumed in many forms including grain, vegetable and fodder. A variety of nutrients are present in the seeds, including carbohydrates, fats, protein, vitamins, minerals, and also some secondary metabolites. Pigeonpea exhibited various ethnomedicinal and pharmacological properties, and it has a long history of ethnobotanical use. Conventional breeding programs are utilized to develop nutritionally improved cultivars, although the success of such a program is very slow due to restricted gene pool and linkage drag. The exploitation of breeding-based approaches along with supportive interdisciplinary research and development have been utilized for biofortified pigeonpea development. Some transgenic approaches were also undertaken for nutritional improvement and antibody production. Further improvement in those approaches and genomic technologies will enhance the nutritional quality of pigeonpea.

Keywords

Genomics-assisted breeding · Health related traits · Molecular markers · Nutraceuticals · Quantitative trait loci · Whole genome sequence

1 Introduction

Pulses hold a salient position in Indian Agriculture. India tops the list for being the largest producer and consumer of pulses in the world, contributing about 25% to the global pulse or grain legumes production (Saxena et al. 2019). One such important grain legume, which has originated from the Indian subcontinent, is the pigeonpea (*Cajanus cajan* (L.) Millspaugh). Predominantly grown in rain-fed conditions, pigeonpea is a considerable source of protein to rural and urban households in Asia and Africa. It augments and enhances the soil through symbiotic nitrogen fixation and revitalizes the soil by recycling of soil nutrients, releasing soil-bound phosphorus, and addition of organic matter (Pahwa et al. 2013). Moreover, it is a great source for additional nitrogen supply to the subsequent crops. According to studies, pigeonpea releases roughly 40 kg/ha of residual nitrogen in the crop fields. All these properties cooperatively make pigeonpea a supreme crop for sustainable agriculture, around the equatorial regions of India. About three-fourths of the total

Indian production of pigeonpea is derived from Gujarat, Karnataka, Maharashtra, Madhya Pradesh, Andhra Pradesh, and Uttar Pradesh. In Barbados, pigeons were fed with these pigeonpea seeds grown on barren lands; this justifies the name. Being a short-day plant, it has a longevity of 3–4 years. Plough pan is formed below the normal ploughing zone and is a compact soil layer, which reduces the productivity of the land. The long tap roots of pigeonpea are prominently known as "biological plough" because of their ability to break plough pan.

The pigeonpea seeds consist of three structural features – cotyledons, seed coat, and embryo. The embryo is rich in albumin, globulin, and the cotyledons have high carbohydrate content, along with calcium and iron (Figs. 1 and 2). The albumin has affluent number of amino acids rich in sulphur; which encompasses methionine and cystine. Other amino acids like glycine, lysine, alanine, and aspartic acid are also present. Methionine is a limiting essential amino acid and hence, it is a beneficial factor under nutrition. The pigeonpea seed coat majorly contains amino acids like serine, proline, threonine, and glycine (Saxena et al. 2019).

The pigeonpea seeds are an integral part of Indian diet. The dry seeds are dehusked and split into cotyledons which are commonly cooked as "dal." In many Indian states, the green seeds serve as a protein-rich vegetable. To garner highest seed yield and utmost nutritional quality, the green pods must be harvested at an appropriate stage. An inverse relationship was observed between the starch content and the sugar-protein contents. In the developing seeds, there is a drop in the sugar and protein content and a rapid elevation in the starch content whereas, iron, zinc, calcium, magnesium, and copper contents were found to be more or less unchanged during seed development in pigeonpea.

Pigeonpea also holds certain antinutritional factors. Polyphenols such as tannins and phenols, oligosaccharides, lectins, enzyme inhibitors like chymotrypsin and trypsin are some of the above mentioned factors (Toklu et al. 2021). Trypsin and chymotrypsin inhibitors are expressed only in the seeds. Whole seeds without dehulling are also consumed in many countries. Cooking of pigeonpea also plays a significant role which affects its nutritional features. The seeds are large in size, absorb more water, and have high nitrogen content, which makes it a quick cooking dal. Cooking not only enhances the bioavailability of certain nutrients, it also destroys certain antinutritional components. For instance, starch digestibility is improved by cooking whereas there is a drop in the measure of oligosaccharides. Heat destroys thiamine and riboflavin, but niacin content remains unchanged during roasting and cooking of pigeonpea seeds. Methionine and lysine content decreases upon roasting, whereas there are reports on increased methionine upon boiling.

Pigeonpea possesses many herbal properties which are essentially described in folk medicine and used to treat numerous human illnesses (Salehi et al. 2019). Pneumonia, bronchitis, coughs can be cured using floral extracts of pigeonpea. It can also be employed to treat respiratory infections, menstrual distress, and dysentery. Dried seeds have the ability to ease difficulties like headache and vertigo, whereas fresh seeds help to diminish urinary incontinence, as well as other kidney disorders. The seed extracts aid in curing sickle cell anemia, by impeding the

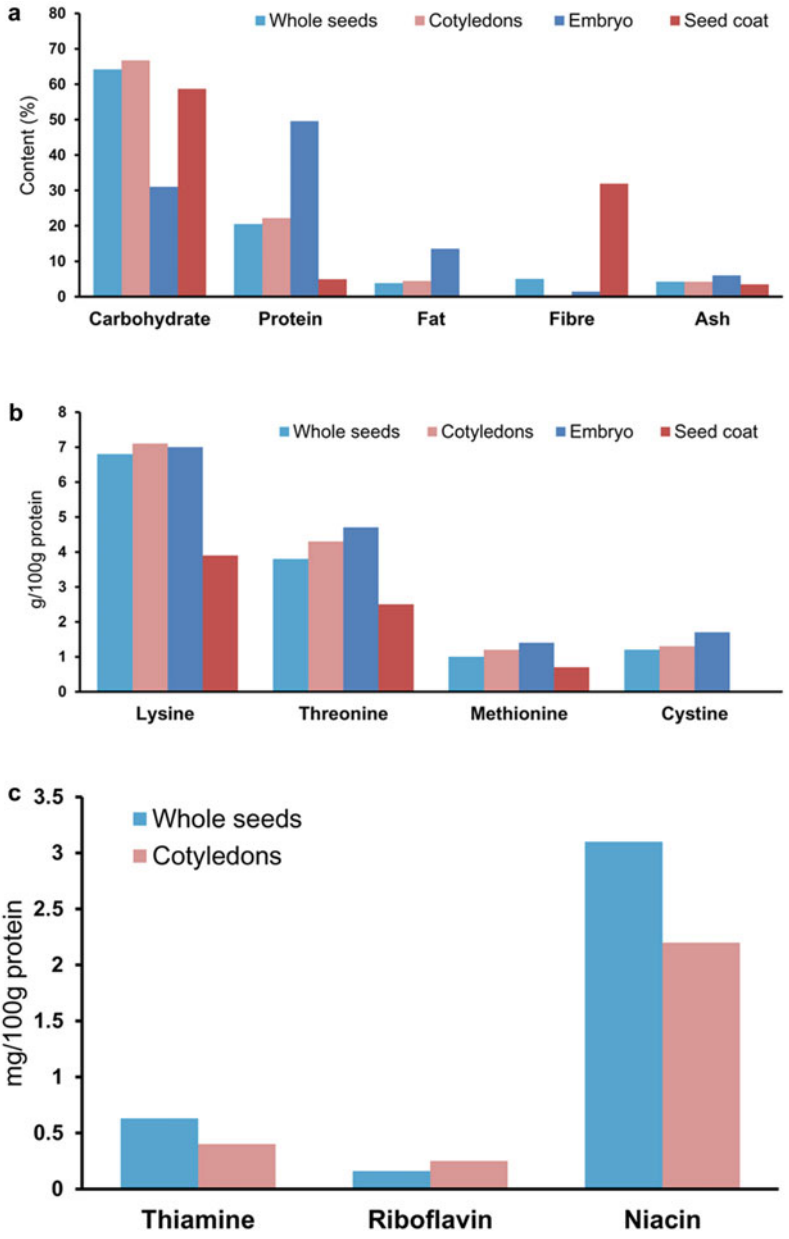


Fig. 1 Overall nutrient composition and their distribution in pigeonpea. (**a**) Nutrients in mature pigeonpea. (**b**) Major amino acids in mature pigeonpea. (**c**) Vitamins in mature pigeonpea. (**d**) Minerals and trace elements. (**e**) Protein fractions in dry pigeonpea seeds (Saxena et al. 2002)

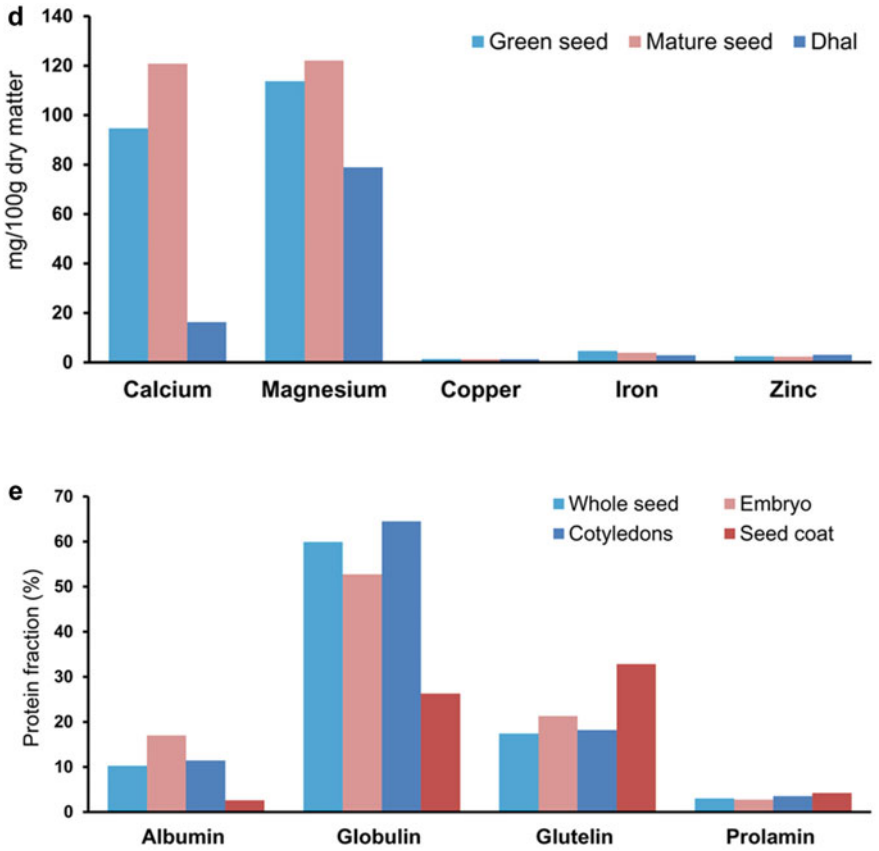


Fig. 1 (continued)

sickling of erythrocytes. According to some reports, dried pigeonpea roots could be used as anthelmintic, sedative, vulnerary, expectorant, and alexiteric.

2 Limitations in Conventional Breeding and Rationale of Nutritional Genomics

Improving the yield quantity, nutritional quality, and maintenance of genotype stability are the primary approaches to fulfil the demands of the population. Conventional breeding practices coupled with genomics-based selection approaches need to be employed to fight the threats offered by climate change and increasing population (Singh et al. 2020).

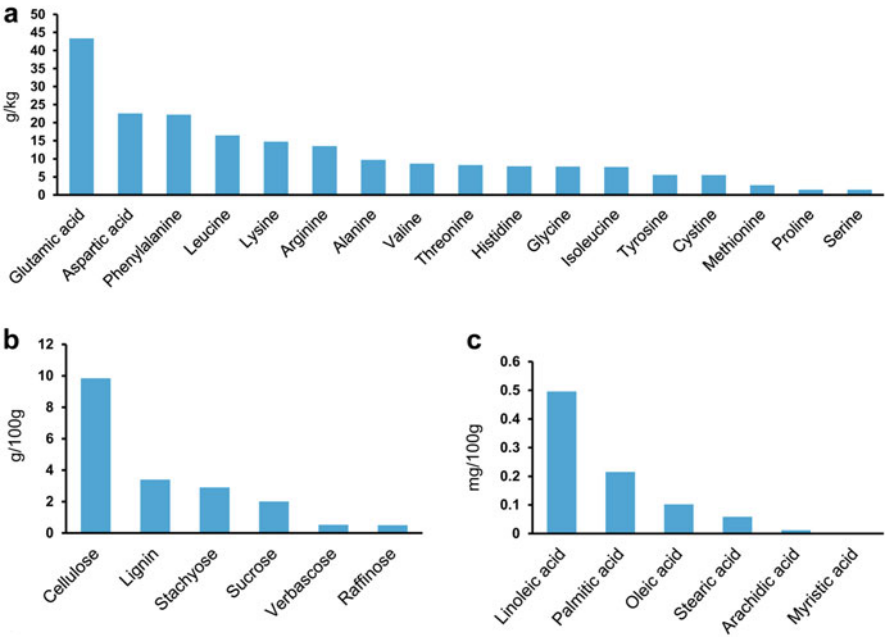


Fig. 2 Detailed nutrient composition of pigeonpea. (**a**) Amino acid composition (Ade-Omowaye et al. 2015). (**b**) Carbohydrate profile (Apata 2008). (**c**) Fatty acid profile (Ade-Omowaye et al. 2015)

Traditional plant breeding methods include the recognition and development of improved parental lines that has quality nutrient content, hybridization with elite genotypes, followed by selection of hybrids over a number of generations to get commercially established cultivars showing required nutritional properties. Additional considerations include quantitative trait complexity and the difficulty of selection of desirable trait because of low heritability. As a result, traditional methods take longer to grow a new and improved variety. Advancement in omics techniques in combination with breeding programs have a lot of potential to contribute for nutritional quality improvement in pigeonpea (Singh et al. 2020). Some of the constraints related to nutritional improvement of pigeonpea are detailed in the next few paragraphs.

Limited diversity within the basic pool of genes was revealed by a polymorphism study of sampled *Cajanus* accessions. Breeders have no choice but to use species and sub-species from secondary, tertiary, and quaternary gene pools through conventional and marker assisted selection techniques. Despite of vast genetic diversity of wild relatives, there is limitation of incorporation of them in breeding program because of lack of accurate information on the availability of desirable features and the necessity for extensive research whenever they are used. Poor agronomic traits in combination with partial characterization of relatively few wild relatives are responsible for lag in genetic improvement of pigeonpea (Saxena et al. 2014).

Pigeonpea is a short-day plant (Vales et al. 2012). A pivotal regulator of flower induction is the interaction of the photoperiod with day and night temperature. Hence, beyond 30° northern and southern latitudes, the cultivation of pigeonpea is restricted. (Saxena 2008). There is an inverse correlation between earliness and photosensitivity which confirms limited success of breeding programs in photo-insensitive and late maturing cultivars. Low-temperature in combination with photoperiod and sensitivity limit the cultivation of this crop in higher altitudes and latitudes (Vales et al. 2012). This is restricting the use of pigeonpea in alternative cropping systems (Vales et al. 2012).

The transfer of the genes of interest into the elite cultivar is highly interfered by the association of unwanted phenotypes with certain nutritional traits. As an example, transferring the genes involved in high protein accumulation was tried from *C. scarabaeoides* and *C. albicans* to the cultivars of pigeonpea. The selection of the desired genotype, high in productivity and protein yield, was obtained only after some 12–14 generations (Saxena and Sawargaonkar 2015).

3 Medicinal Properties of Pigeonpea

Pigeonpea had been used extensively in traditional medicine. In addition, the plant has also exhibited a wide array of pharmacological properties. This section will describe the various ethnomedicinal and pharmacological properties of pigeonpea along with a brief illustration about the selected chemical constituents present in the plant.

3.1 Ethnomedicinal Uses

The *Garo* tribal community of Netrakona district of Bangladesh uses pigeonpea as a remedy for diabetes. The seed paste of this plant is used as a stimulant while the leaf juice is used for the treatment of diabetes (Rahmatullah et al. 2009). In Trinidad and Tobago, the plant is used to treat food poisoning and is considered as colic. It is also used to treat constipation (Lans 2007). In Cote D'Ivoire, extraction from the leaves and stem are utilized for the treating of anemia, skin disease, and wounds (Koné et al. 2011). In Benin, the similar preparation is used for the treatment of candidiasis (Fanou et al. 2020). The local communities of south western Uganda use the juice of the leaves for the treatment of ear disease (Gumisiriza et al. 2019). In south west Nigeria, the leaves of the plants are used for treating malaria (Olorunnisola et al. 2013).

3.2 Active Principles of Pigeonpea

Chemical analysis revealed high quantities of flavonoids and stilbenes in the leaves of pigeonpea. Saponins, a significant quantity of tannins, and modest amounts of reducing sugars, resin, and terpenoids were also reported from the plant (Pal et al. 2011). Pigeonpea flavonoids can be found in a variety of plant organs. There are 27 flavonoids

present. Among them flavones, isoflavones, and flavonols have been noticed in six, eight, and four numbers respectively. Besides them two anthocyanins and several flavanones, isoflavanones are also recorded, along with a solitary chalcone (Nix et al. 2015). Table 1 illustrates the selected flavonoids present in the plant. Apart from these, the plant contains stilbenes in the form of longistylene A (Wu et al. 2020) and longistylene C (Wang et al. 2011). *Cajanus* lactone and cajaninstilbene acid (Wu et al. 2009) along with pinostrobin have also been reported from the leaves of the plant (Patel and Bhutani 2014).

3.3 Pharmacological Uses of Pigeonpea

Since ancient times, different portions of pigeonpea have been used for their biological activity, and some of them have experimental grounds for acceptability. Aside from their use in traditional medicine, there have been various studies on pigeonpea's biological and pharmacological properties (Table 2).

3.3.1 Antibacterial Activity

The antibacterial activity of pigeonpea has been explored in a number of studies. In one experiment it was shown that the ethyl acetate leaf extraction contains naringenin that inhibited growth of *Salmonella typhi* and Staphylococcus *aureus* indicating its potential in the treatment of typhoid (Agus et al. 2017). It was shown that organic solvents extractions and water extracts were inhibiting *Escherichia coli, Staphylococcus aureus* growth, whereas *Klebsiella pneumonae* was inhibited by the extracts of organic solvents only. In addition, the minimum concentration of extract to inhibit *E. coli* was recorded as 0.125–0.25 mg/ml; to inhibit *S. aureus* it was found to be 0.125 mg/ml and that of *Salmonella typhi* was to be 0.0325–0.0625 mg/ml (Okigbo and Omodamiro 2007).

3.3.2 Antifungal Activity

Antifungal activity of the plant was evaluated using ethanolic extract of leaf and root. It was observed that extracts inhibited growth of *Candida albicans* and *Candida tropicalis*. Tannins, flavonoids, and alkaloids in extracts from both organs was discovered to have clinically significant antifungal activity (Brito et al. 2012).

3.3.3 Antiviral Activity

One study looked at the activity of water and ethanolic extracts against the measles virus as well as its toxic effect to embryonated chicken eggs. The in vivo assay using stem extraction in water provided a Log(2) titre of 0.1, and when the assay was done in vitro, a 100% suppression of cytopathic effect was observed in cell lines of Hep-2. Hemagglutination titration revealed a decrease in viral content ($p = 0.05$) at all concentrations of the extracts (Nwodo et al. 2011).

Table 1 Selected flavonoids isolated from pigeonpea

Types	Name	IUPAC name	Source plant organ	References
Flavones	Apigenin	5,7,4′-trihydroxyflavone	Leaves	Fu et al. 2008
	Luteolin	5,7, 3′,4′-tetrahydroxyflavone	Leaves	
	Vitexin	Apigenin 8-C-glucoside	Leaves	Fu et al. 2007
	Isovitexin	Apigenin 6-C-glucoside	Leaves	
	Orientin	Luteolin 8-C-glucoside	Leaves	Wei et al. 2013
Isoflavones	Biochanin A	5,7-Dihydroxy-4′-methoxyisoflavone	Leaves and roots	Duker-Eshun et al. 2004
	Cajanin	5, 2′,4′-Trihydroxy-7-methoxyisoflavone	Seed and etiolated stems	Dahiya 1987
	Genistein	5,7,4′-Trihydroxyisoflavone	Roots/root, bark, and etiolated stems	Duker-Eshun et al. 2004
	2′-Hydroxygenistein	5,7,2′,4′-Tetrahydroxyisoflavone	Roots/root, bark, and etiolated stems	Duker-Eshun et al. 2004
Flavonols	Quercetin	3,5,7,3′,4′-Pentahydroxyflavone	Leaves	Zu et al. 2006
	Isoquercitrin	Quercetin 3-β-D-glucoside	Pod surface	Green et al. 2003
	Isorhamnetin	3′-Methoxyquercetin	Leaves	Zu et al. 2006
Flavanones	Naringenin	5,7,4′-Trihydroxyflavanone	Leaves	Wei et al. 2013
	Pinostrobin	5-Hydroxy-7-methoxyflavanone	Leaves	Wei et al. 2013
Isoflavanone	Cajanol	5,4′-Dihydroxy-7,2′-dimethoxyisoflavanone	Roots	Luo et al. 2010
	Cajanone		Roots	Dahiya 1991
Chalcone	Pinostrobin	2′,6′-Dihydroxy-4′-methoxychalcone	Leaves	Patel and Bhutani 2014

Table 2 Pharmacological activities of pigeonpea

S. no.	Pharmacological activity	Parts	Form used	Active principle involved	References
1.	Antibacterial activity	Leaf	Ethyl acetate fraction of leaf extract	Naringenin	Agus et al. 2017
2.		Leaf	Petroleum ether, ethanol, and chloroform/ methanol mixture extracts (organic) Aqueous extract		Okigbo and Omodamiro 2007
3.	Antifungal activity	Leaf Root	Ethanolic extract of leaf and root		Brito et al. 2012
4.	Antiviral activity	Leaf Stem Root	Hot water and ethanol extract of leaf, stem, and root		Nwodo et al. 2011
5.	Antimalarial activity	Leaf	Methanol extract Column chromatographic technique with organic solvent systems used to isolate compound	Cajachalcone	Ajaiyeoba et al. 2013
6.		Root	Ethanolic extract of roots	Longistylin A and C, and betulinic acid	Duker-Eshun et al. 2004
7.	Antidiabetic activity	Root	Methanolic extract of roots		Nahar et al. 2014
8.		Leaf	Methnolic extract		Ezike et al. 2010
9.	Hypocholesterolemic effect	Leaf	Ethanolic extract followed by extraction with hexane and dichloro ethane	Cajanin, Longistylin C, and Longistylin A	Luo et al. 2008
10.	Hypolipidemic effect	Leaf	Methanolic extract		Akinloye and Solanke 2011

11.	Neuractive activity	Leaf	Ethanolic extract followed by partitioning and column chromatography using organic solvents	Pinostrobin	Nicholson et al. 2010
12.	Antihelminthic activity	Leaf	Extraction with petroleum ether, ethyl acetate, ethanol, and water		Khan et al. 2015
13.	Hepatoprotective	Leaf	Ethanolic extract		Iweala et al. 2019
14.	Anti-inflammatory activity	Root	Hot water and ethanolic extract extract		Vo et al. 2020
15.	Anti-inflammatory and antinociceptive activities	Seed	Hexane extracts	Quercetin-3-*O*-β-D-glucopyranoside, Orientin, Vitexin, Quercetin, Luteolin, Apigenin, Isorhamnetin	Hassan et al. 2016
16.	Anticancer activity	Root	Pure compound	Cajanol	Luo et al. 2010
17.		Root	Pure compound	Cajanin	Fu et al. 2015
18.		Stem, roots	Aqueous extract		Teixeira et al. 2021
19.	Antioxidant activity	Leaf	Aqueous extract Ethanol extract Petroleum ether extract Ethyl acetate fraction *n*-Butanol fraction	Cajaninstilbene acid Pinostrobin Vitexin Orientin	Wu et al. 2009

3.3.4 Antimalarial Activity

Antimalarial activity of the plant was determined in vitro utilizing *Plasmodium falciparum* (K1) which is a multiresistant strain. This variant was used in the parasite lactate dehydrogenase assay employing bioassay-fractionation of the pigeonpea leaf extraction in methanol. Various chromatographic techniques were used to isolate the compound, and spectroscopy was used to determine its structure. The physiologically active ingredient from the ethyl acetate fraction was identified as a cajachalcone also known as 2′,6′-dihydroxy-4-methoxy chalcone. The IC_{50} of cajachalcone was 2.0 μg/ml (7.4 μM). *Plasmodium falciparum* was inhibited by the extracts containing active principle (Ajaiyeoba et al. 2013). In another study, it was observed that in vitro assays performed with the *Plasmodium falciparum* strain 3D7 that shows chloroquine-sensitivity was moderately strong for various compounds like betulinic acid, longistylin A and C, stilbenes (Duker-Eshun et al. 2004).

3.3.5 Antidiabetic Activity

The antidiabetic activity of the methanolic root extract was monitored using alloxan-applied mice with diabetes for 5 days. This indicated that upon oral ingestion of extracts of plant at various doses of body weight (200–400 mg/kg), there was a significant reduction in serum fasting glucose in diabetic mice induced with alloxan (Nahar et al. 2014). Some studies demonstrated that when alloxan applied mice, showing diabetes, were administered with 400–600 mg/kg of methanolic extract, the fasting blood sugar reduced with maximum effect between 4 and 6 h (Ezike et al. 2010).

3.3.6 Hypocholesterolemic Effect

Hypocholesterolemic effect of the leaf extraction of pigeonpea was evaluated on diet-induced hypercholesterolemic mice. Excessive levels of serum and cholesterol from liver were significantly lessened by the 200 mg/kg plant extract after 4 weeks pretreatment, comparing to the model, by nearly 31% and 23% ($p = 0.01$), respectively. The proportions of serum and liver triglycerides were also minimized by 23% and 14%, respectively. During this time, LDL cholesterol from serum reduced by almost 53% ($p = 0.01$), whereas superoxide dismutase activity from serum rose by nearly 21%. The body weight and atherogenic index were both significantly lowered. mRNA transcript accumulation of HMG-CoA reductase, LDL-receptor, and CYP7A1were dramatically increased in mice given 200 mg/kg/day of plant extract, but the hypercholesterolemic diet repressed those expressions (Luo et al. 2008).

3.3.7 Hypolipidemic Effect

Methanolic extraction from leaves of the plant was tested for its hypolipidemic effect. The result showed a significant ($p = 0.05$) reduction in cholesterol, serum triglyceride, HDL, LDL, cholesterol, and blood glucose. The extract also reduced the functionality of aspartate transaminase and alanine transaminase along with reduction in levels of creatinine, urea and malondialdehyde levels in alloxan induced hyperglycemic mice (Akinloye and Solanke 2011).

3.3.8 Neuroactive Activity

Pinostrobin, from pigeonpea, was studied in vitro for its neuroactive characteristics and was found to inhibit voltage-gated sodium channels ($IC_{50} = 23$ μM). This study was based on the previously known background about pinostrobin, which has the capacity to reduce the depolarization effects of a certain selective activator of sodium channels called veratridine, in the brain synaptonemal complex of mice. This compound had nil effect on synaptoneurosomes resting membrane-potential. Pinostrobin's pharmacological profile is similar to that of depressive medications that block sodium channels (Nicholson et al. 2010).

3.3.9 Anthelminthic Activity

Antihelminthic activity was assessed using the ethanolic and aqueous extract of the pigeonpea. The results suggest that, aqueous extraction has anthelmintic action for paralyzing and killing Indian earthworm *Pheritima posthuma* for a long period at 5 mg concentration, whereas the ethanolic extract has paralysis and death in a short time at the same dosage (Khan et al. 2015).

3.3.10 Hepatoprotective Activity

The hepatoprotective activity of the plant was studied with respect to hepatotoxicity in male wistar rats. *N*-Nitrosodiethylamine (NDEA) induced hepatotoxicity which was reversed by the ethanolic extract of the leaf of the plant. The results indicated that pigeonpea-treated groups had considerably ($p = 0.05$) lower alanine and aspartate aminotransferases levels and significantly ($p = 0.05$) higher glutathione *S*-transferase, superoxide dismutase, glutathione, albumin, and catalase levels (Iweala et al. 2019).

3.3.11 Anti-inflammatory Activity

The anti-inflammatory activity of pigeonpea was evaluated in an in vitro experiment using RAW 264.7 cells. The results confirmed that 95% ethanolic extract of the roots dramatically reduced intracellular reactive oxygen species and increased superoxide dismutase and catalase activity. EECR95 induced nuclear factor (NF) erythroid 2-related factor 2/antioxidant protein heme oxygenase-1 and hindered nuclear factor kappa B (NF-B) signaling pathways, resulting in antioxidant and anti-inflammatory properties, according to mechanism studies (Vo et al. 2020). In another experiment, albino rats were used as experimentation models to study the anti-inflammatory and antinociceptive activities of the plant seeds. The results indicated that in hexane extract of seeds, twenty-one unsaponifiable chemicals (including various phytols, stigmasterol, 2,6-di-(t-butyl)-4-hydroxy-4-methyl-2,5-cyclohexadiene-1-one, campesterol, and sitosterol) as well as fatty acids described mostly as palmitic acids and 9,12-octadecadienoic, almost 12 in numbers were found. Quercetin, Orientin, Luteolin, Quercetin-3-*O*-D-Glucopyranoside, Vitexin, Apigenin, and Isorhamnetin are all found in the n-butanolic extraction part. Three hours after carrageenan challenge, the hexane extract (200 and 400 mg/kg) reduced carrageenan induced inflammatory effects by a significant 85% and 95%, respectively. This was associated by a reduction in TNF- and IL-6 levels of 11% and 20%, 8% and 13%, respectively, as well as a significant reduction in IgG serum quantity. In

addition, hexane fraction (200 and 400 mg/kg) reduced writings by 61 and 83%, respectively (Hassan et al. 2016).

3.3.12 Anticancer Activity

The anticancer activity of cajanol, an isoflavanone derived from pigeonpea roots, was noted in a study using breast cancer cell lines from human (MCF-7). Cajanol suppressed MCF-7 cell growth depending upon dose- and time-specificity. After 24 h of treatment, the IC_{50} value was 83.42 μM, reached 58.32 μM after 48 h, and reduced to 54.05 μM after 72 h. Cajanol used a ROS-mediated mitochondria-dependent route to inhibit the cell cycle in the G2 and M stage and cause programmed cell death. Cajanol blocked the expression of Bcl-2 expression and elevated expression of the Bax gene, which led to the rupture of the outer mitochondrial membrane and resulted in cytochrome c liberation, as experimented through Western blot. The induction of the caspase-9 and caspase-3 cascades was linked to mitochondrial cytochrome c release, while active-caspase-3 was engaged in PARP cleavage (Luo et al. 2010). Another research showed cajanin stilbene acid obtained from the plant were investigated for its anticancer properties. Cajanin caused apoptosis and G2/M inhibition in a concentration-specified manner. Matrix Metalloproteinases was degraded, Bax level was increased, Bcl-2 was decreased, and caspase-3 was induced. BRCA-specific DNA impairment responsive pathways as well as cell cycle-controlling chromosome replicative pathways were both impacted by cajanin stilbene acid, according to microarray profiling (Fu et al. 2015). Other study indicated that the fractions of stem and root extracts inhibited melanoma proteases and generated cellular toxicity in SK-MEL-28 cells, cultured in vitro (Teixeira et al. 2021).

3.3.13 Antioxidant Activity

In a study, the antioxidative nature of pigeonpea from aqueous and ethanolic leaf extracts, as well as ethyl acetate, n-butanol, petroleum ether, and water fractions, as well as the four main compounds separated from the ethanol extract, namely pinostrobin, cajaninstilbene acid, orientin, and vitexin were investigated by a DPPH radical-scavenging assay. An IC_{50} value of 194.98 μg/ml, the ethyl acetate fraction had the highest scavenging power among the four fractions. Pinostrobin and vitexin were shown to have less effective radical-scavenging powers than cajaninstilbene acid (302.12 μg/ml) and orientin (316.21 μg/ml). The inhibition ratio (%) of the ethyl acetate fraction (94.13% ± 3.41%) was found to be the greatest in the beta-carotene-linoleic acid test, practically matching the inhibitory capability of the positive control BHT (93.89% ± 1.45%) at 4 mg/ml. When compared, cajaninstilbene (321.53 μg/ml) and orientin (444.61 μg/ml) had moderate antioxidant effects, while pinostrobin and vitexin both exhibited antioxidant activities at greater than 500 μg/ml (Wu et al. 2009).

4 Genetic Resources of Health-Related (HR) Genes

A large number of genetic resource accumulations, including genetic maps, molecular markers, whole-genome resequencing (WGRS) data, transcriptome assemblies, a reference genome sequence (Fig. 3) (Varshney et al. 2012) from multiple cultivars,

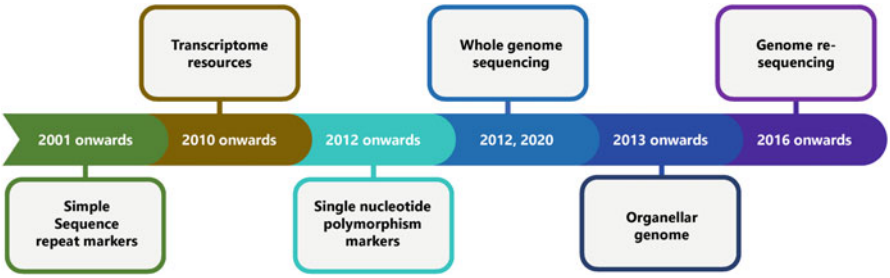


Fig. 3 Timeline of major genomic approaches adopted in pigeonpea

have become available in pigeonpea (Kumar et al. 2016; Varshney et al. 2017). These resources have aided in the creation of high-resolution genetic maps as well as efficient and expeditious genetic analysis of quantitative trait loci (QTLs) and genes regulating important nutritional traits in pigeonpea (Saxena et al. 2012).

Pigeonpea is an essential food source with amino acid rich plant protein for more than a billion people worldwide. However, genetic improvement for seed protein content (SPC) in the crop has acquired little concern in the past. The use of genomics-assisted breeding could aid in the acceleration of SPC genetic gain. Four genotypes of pigeonpea were taken for whole-genome resequencing data to recognize sequence-based markers and associated possible SPC genes (Obala et al. 2019). One hundred and eight sequence variations obtained from 57 genes were recognized by combining a common variant sieving methodology on already procured WGRS data with the gene functioning data concerning SPC. Subsequently, 17 of the 30 sequence variants when transformed into CAPS/dCAPS markers showed significant polymorphic traits between genotypes of low and high SPC. A significant ($p = 0.05$) co-segregation of 4 of the CAPS/dCAPS markers was observed with SPC when 16 polymorphic CAPS/dCAPS markers were tested on F_2 generation which is a cross of ICP 5529 and ICP 11605, former with high SCP and the latter with low SCP. In summary, mutations in four gene sequences gave rise to four markers and were suggested to be helpful in pigeonpea crop improvement programmes for enhancing/regulating SPC (Obala et al. 2019).

5 Classical Genetics and Traditional Breeding for HR Traits

Over the last two decades, many attempts have been made to create high-yielding cultivars by traditional breeding methods and advancements in biotechnology. These investigations have given information and understanding for creating superior pigeonpea varieties with many agronomically important quality characters and show great yield potential even in challenging agro-climatic settings. New cultivars with better nutritional content and boosting production potential have already been created using traditional plant-breeding techniques. To develop genotypes with the required nutritionally rich and agronomically superior features, classical plant breeding requires identifying and developing parental lines showing enhanced nutrition-rich content, crossing the latter with elite germplasm, and selection of the

segregating population for some generations (Pfeiffer and McClafferty 2007). Thus, it pertains to a much-extended time to procure a novel or better variety. The complications at genetic level of quantitative traits and low heritability are some bottlenecks that pose challenges for selecting superiors.

Due to a number of specific features, breeding of pigeonpea has proven to be more difficult than breeding other edible legumes. Pigeonpea is often cross-pollinated crop. Insect-aided natural outcrossing rates of 20–70% in pigeonpea, have restricted the application of effective selection and mating methods are available in self-pollinating species (Saxena and Sharma 1990). This crop's yield potential has gradually increased due to the employment of extensive hybridization, pure line breeding, population breeding along with mutation breeding hence create new pigeonpea varieties. Two genetic male-sterility (GMS) systems were found in pigeonpea to help with this bottleneck (Reddy et al. 1979). The GMS-based hybrids had a yield which was 30% more than that of nonhybrids but did not prove to be commercially viable because of its exorbitant production cost.

The alternative and more effective cytoplasmic-genetic male-sterility (CGMS) approach was created in response to the yield-jump seen in the GMS hybrids (Saxena and Kumar 2003). In 2004, India had its first cytoplasmic male sterility (CMS)-based hybrid GTH-1 available from ICRISAT's hybrid development programme in partnership with its partners. Furthermore, another CMS-based pigeonpea hybrid, ICPH 2671, was created in 2005 at ICRISAT utilizing *C. cajanifolius* (A4 cytoplasm) and has since been commercially available by Pravardhan Seeds under the name "Pushkal" for cultivation in various Indian states, including Maharashtra, Madhya Pradesh, Karnataka, and Andhra Pradesh. The expanded area cultivating pigeonpea hybrids is projected to result in higher crop yield and satisfying returns for farmers and pigeonpea production in a sustainable manner was possible. This will again be made feasible by ongoing attempts to breed resistance to biotic and abiotic challenges.

Besides breeding for yield, breeding for nutrition has always been the focus of pigeonpea breeders. Despite pigeonpea being the household dal, consuming every single day, the average protein requirement of an Indian adult is not met. Hence, a breeding programme was initiated back in 1982 at ICRISAT. ICRISAT's genebank houses 13,632 germplasm which has a protein range from 9% to 30% (Varshney et al. 2012). Protein content in pigeonpea is controlled by additive genetic action. Based on available information from the genebank, wild progenitors *C. scarabaeoids* (28.4%), *C. sericeous* (29.4%), and *C. albicans* (30.5%) were utilized to develop new protein lines. Accordingly, newly bred lines, called high protein lines (HPL) reported protein content up to 32%. These lines are in preliminary yield testing stage and serve as a donor for high protein trait in a breeding program. This twenty-first century has greater innovation in terms of protein. Protein based markets are worth USD 38 billion (2019) and is expected to grow at a rate of 9.1% from 2020 to 2027. Increasing traction towards plant-based protein (either as protein isolate or protein concentrates) is a greater opportunity for paradigm shift in nutritional breeding. Utilization of indigenous crops for protein source has been the current focus in Indian protein market. "Smart Protein" is a budding concept, pulses including pigeonpea is a part of this initiative. Harnessing the protein content of indigenous crops to be used as alternative protein

source without burdening the environment is the aim. With nonmeat, vegan, dairy-free, vegetarian, and ethical food systems in rise "smart protein" will be the future.

Next nutritive trait is Fe and Zn. The recommended daily allowance (R.D.A.) of Fe for a child and an adult in India is 13 and 17 mg per-day, respectively. Whereas the R.D.A. of per-day Zn for a child and an adult is 7 and 12 mg. Nevertheless, a food proportion of 7 g a day per person in India, imparts a daily per capita iron intake of 14.93 mg, which is much lesser than R.D.A. With this backdrop, a baseline study of genetic variability was taken for Fe and Zn content in pigeonpea at ICRISAT. Accordingly, a range of 24.91–44.65 mg/kg seed for Fe content and 26.08–47.80 mg/kg seed for Zn content was noted. Both wet methods, as well as Energy-dispersive X-ray fluorescence technique, were used to calibrate and estimate whole seed Fe and Zn content. A breeding programme is halfway in fortifying for Zn and Fe in pigeonpea. Marker-assisted backcrossing is effectively carried out for forwarding the generation.

Recent development of early and photo-insensitive pigeonpea lines coupled with rapid-generation turnover methods has helped in fast-forwarding the generation. Interestingly, early genotypes are high in nutritional traits and is a win-win situation for introgression and generation advancement. Unlike the 1990s, three cropping seasons with year-round breeding can now be done. Conventional breeding coupled with genomic selections has increased the selection efficiency. Reduction in time taken for completion of a cropping season has increased the genetic gains in pigeonpea.

6 Genetic Diversity with Regard to HR Traits

Molecular markers play a pivotal role in genetic improvement program of any crop. These are used both in the genetic diversity assessments as well as trait-specific molecular mapping. Various kinds of molecular markers have been adopted in pigeonpea also including first generation restriction fragment length polymorphism (RFLP), and subsequently, random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR), and latest single-nucleotide polymorphism (SNP) (Saxena et al. 2014; Pazhamala et al. 2015) markers. Amongst these, SNP markers stand for ideal DNA marker owing to their higher abundance throughout the genome and high throughput estimation procedure, apart from other advantages of a codominant marker.

WGRS was given about 292 accessions to track the genetic diversity of pigeonpea. This included wild species, landraces, and breeding lines, yielding a total count of 17.2 million variations (Varshney et al. 2017). To discover how several candidate genes were related to agronomically significant variables, a GWAS was conducted. Sequence similarities exist between the genes functionally described in other plants for flowering time control, seed development, and pod dehiscence and the candidate genes for these features in pigeonpea. These polymorphic locations will help create high-density SNP arrays, genotyping of various mapping populations to create genetic maps, and identify the genomic areas underlying significant agronomic features. A total of 932 markers were used to create a condensed intraspecific pigeonpea linkage map, covering an overall adjusted map

length of 1411.83 cM to enhance chromosomal anchoring and to map the genes linked to useful agricultural traits. It contains 65 SSR marker loci, 319 RAD-SNPs, and 547 bead-array SNPs (Arora et al. 2017). The genetic advancement of pigeonpea could be sped up with the help of this information. Recently, two high-density Affymetrix Axiom genotyping chips have been created in pigeonpea to accelerate the genetic gain. A 56 K *Cajanus* SNP chip has been created to study the genetic variation across 103 pigeonpea lines (Saxena et al. 2018).

7 Molecular Mapping of HR Genes and QTLs

High-throughput genotyping applications have caused drastic improvements in the density of markers which were used to generate genetic maps of pigeonpea. These have been adopted in pigeonpea, too, for the last two decades. Several genotyping programs targeting the F_2 populations have resulted in high-density genomic maps to date (Arora et al. 2017; Saxena et al. 2017; Yadav et al. 2019). Such genetic resources were crucial to dissect the genomic design of agronomic traits in pigeonpea, including its nutritional appearances. Fine mapping of QTLs responsible for nutritive properties of pigeonpea is essentially required to generate superior cultivars/genotypes with potential well-being properties (Fig. 4).

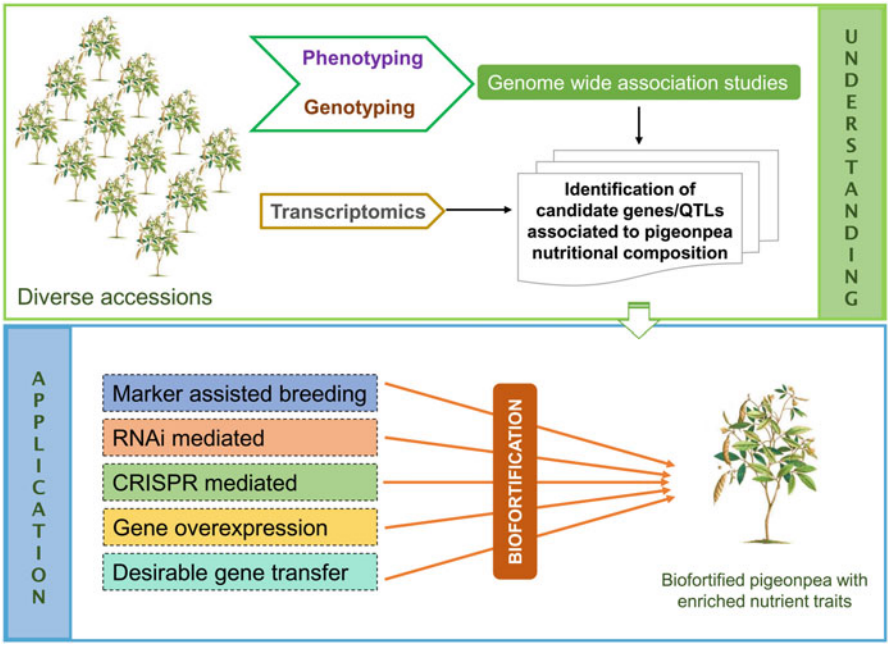


Fig. 4 Overview of concurrent genomic technologies for designing biofortification of pigeonpea

8 Marker-Assisted Breeding for HR Traits

In recent times, the availability of convenient library preparation methods and greater multiplexing capacity has facilitated the genotyping-by-sequencing (GBS) approach as a promising tool for the simultaneous discovery and characterization of numerous SNPs (Saxena et al. 2017). Whole-genome resequencing (WGRS) has become the latest high-throughput option for determining genetic variation and trait-linked marker discovery. Accordingly, an SNP array has been developed by resequencing diverse germplasm of pigeonpea with as many as 56,512 unique informative sequence variations (Saxena et al. 2018). Furthermore, identifying key agronomic traits associated with 1554 SNPs and 385 insertion/deletion (InDel) markers potentially enriched the genomic resource in pigeonpea toward marker-assisted selection. The WGRS-based first-generation HapMap of pigeonpea unveiled 5.5 million genome-wide variants (4.6 million SNPs and 0.7 million InDels) (Kumar et al. 2016). Using a different whole-genome resequencing method, candidate gene sequence-based markers in relation to seed protein content were recognized, using four pigeonpea genotypes (Obala et al. 2019). The first-generation HapMap in *Cajanus* spp. was created using the whole-genome resequencing (WGRS) method to develop genetic resources. In a panel comprising of 20 *Cajanus* spp., including 2 wild and 18 cultivated species, there are 5,465,676 genome-wide variants, comprising 4,686,422 SNPs and 779,254 InDels. These sequence variations make mapping the genomic areas underlying fundamental features possible.

9 Map-Based Cloning of HR Genes/QTLs

Pigeonpeas have a protein level of about 21%. However, because they contain less lysine than other legumes, they have poor nutritional value. Dihydrodipicolinate Synthase, or DHDPS, is a crucial regulator of lysine biosynthesis. The DHDPS genes is inactivated by even trace amounts of lysine via a feedback mechanism, as a result pigeonpea exhibits low levels of lysine. Hence, the pigeonpea was transformed with the mutant DHDPS gene (dhdps-r1 from *Nicotiana sylvestris*), since it is no longer responsive to the feedback inhibition by lysine. DHDPS activity was two to six times higher in transgenic pigeonpea, resulting in an 8.5-fold increase in the amount of free lysine in the seeds (Thu et al. 2007). Additionally, pigeonpea has been utilized in the creation of edible vaccinations. With a transformation efficiency of roughly 67%, the Rinderpest virus's haemagglutinin protein antigen was successfully produced in pigeonpea (Satyavathi et al. 2003). An Indian isolate of the Peste des Petits Ruminants (PPR) virus's hemagglutinin-neuraminidase gene (HN) has also been successfully converted and expressed in transgenic pigeonpea. Neuraminidase activity showed that HN protein was physiologically active in transgenic pigeonpea (Prasad et al. 2004).

10 Genomics-Aided Breeding for HR Traits

Conventionally identified QTLs controlling key agronomic traits in pigeonpea available so far (Bohra et al. 2019; Varshney et al. 2013) are inconvenient due to time challenges, cost, and labor faced by those low-throughput marker systems. The pitfalls of conventional marker systems can be overcome by employing high-density genome-wide marker systems. Genome-wide association study (GWAS) is one of the approaches that address the concern of low precision conventional QTL mapping. Instead, being independent of the biparental mapping population helps better understand the genomic background underlying complex phenotypic traits with higher resolution (Huang and Han 2014; Liu and Yan 2019). Accordingly, association mapping of diverse genotypes came out with the significant number of SSRs and SNPs throughout pigeonpea genome governing multiple traits of interest (Mir et al. 2014; Patil et al. 2017). The breakthrough GWAS of 286 resequenced pigeonpea accessions pinpointed numerous marker trait associations related to domestication and with prospects to breeding (Varshney et al. 2017). Nonetheless, more rigorous genotyping of potential accessions/cultivars and simultaneous high-resolution marker-trait association studies would still be required for the efficient next-generation genomics-assisted breeding programs in pigeonpea.

11 Transgenic Studies

Owing to properties such as rapid growth, lofty protein content, capacity to tolerate drought conditions, and a deep root system, pigeonpea is an economically essential crop. There is a huge breach created between the demand and supply of pigeonpea. This has been caused due to the explosion of population and the interplay of biotic and abiotic stresses affecting the growth of the crop. Biotic factors include certain insect pests, like *Helicoverpa armigera*; and some fungal diseases like Fusarium wilt. Abiotic stresses which lead to a drop in productivity include salinity and water logging. Other factor like extensive use of pesticides and herbicides which decreases soil fertility also effects the production of pigeonpea (Negi et al. 2021). Crop breeding has been the most traditional and well-established method of crop improvement. Plant breeding in pigeonpea is a laborious and time-consuming process. One of its main drawbacks is the restricted genetic diversity that results from gene loss during artificial selection. In order to resolve the issues and increase the pigeonpea production, several biotechnological approaches have been used. One of the most triumphant biotechnological approaches has been transgenic technology which removes the major breeding barriers. The development of transgenic technology has demonstrated remarkable success in pulse crop protection. It has also long-term supported research on the inclusion of agronomically advantageous traits, which improves crops and increases the world's population's access to high nutritious food (Saxena et al. 2016). The effective integration of many foreign genes using recombinant DNA technology has opened up new possibilities for the creation of tolerant pigeonpea cultivars with built-in resilience to survive biotic stress factors (Ghosh et al. 2014a).

The availability of several transformation techniques has facilitated the production of effective transgenic crops in many crop species. Of them, *Agrobacterium tumefaciens*-mediated genetic transformation is the most practical and widely applied method on a variety of plants. Researchers employed genetic transformation technology to improve more than 15 cultivars of the pigeonpea by enhancing nutritional quality or by including resilience against various environmental factors. Transgenic pigeonpea has been developed by incorporating a variety of genes, including cowpea *protease inhibitor* (*CPI*), *Bacillus thuringiensis* endotoxins *cry1A(b), cry1Ab, cry1Aabc, cry1AcF, cry1Ac, cry2Aa, and cry1 E-C,* etc. This has elevated the toxicity against the lepidopteran insects (Nandini et al. 2022).

The antibiotic selection based in vitro tissue culture approaches showed numerous drawbacks despite extensive use, such as after successful transformation, small percentage of totipotent cells were able to survive, the selection pressure lowering the explants' overall capacity for regeneration, and inadequate rooting responses (Ghosh et al. 2014b). In 2008, Ramu et al. first introduced *in planta* transformation method which fully skipped the in vitro co-cultivation and selection process and produce a large number of transgenics. Ghosh et al. (2017) developed a unique shoot grafting technique to develop Cry1Ac and Cry2Aa transgenic pigeonpea lines with steady DNA integration up to the T_2 generation. Furthermore tissue culture independent technique was introduced by Ganguly et al. (2018) as plumular meristem transformation method with increasing transformation frequency and PCR based screening process.

11.1 Transgenic Pigeonpea Development for Biofortification

Various agronomically important genes has been discovered in well-characterized systems like *Arabidopsis*, tobacco, rice, pea, carrot, and other plants, and scientists were working to create transgenic pigeonpea plants that were resistant to biotic, abiotic stresses, and with good agronomic traits. (Banu et al. 2014).

The ability to fix nitrogen in the roots is one of the most significant crop-specific characteristics of pigeonpea. This attribute improves and increases soil fertility. However, due to its high fixation in soil and low mobility, availability of phosphorous is constrained. As an adaptive strategy, plants vary the number of lateral roots, develop excessively root hair, and exude organic acids, particularly citrate to alter the rhizosphere (Shen et al. 2005). In order to refine and upgrade P uptake, Transgenic pigeonpea was created by overexpressing *Daucus carota citrate synthase* (DcCs) gene from carrot (*Daucus carota*), under a constitutive and root specific promoter. In both P deficient and P available situations, transgenic pigeonpea lines overexpressing the DcCs gene demonstrated higher level citrate synthase production and enhanced root growth (Hussain et al. 2016).

Pigeonpea serves as an important source of protein, often high lysine content and complements the protein in cereals. Although during agricultural processing, lysine and tryptophan are lost in large amounts (Singh and Eggum 1984). Additionally, *Dihydrodipicolinate synthase* (*DHDPS*), the main enzyme of lysine biosynthesis pathway is also feedback-inhibited by lysine. Under the control of a phaseolin seed-

specific promoter, a mutant *dhdps-r1* gene from *Nicotiana sylvestris* that expresses a lysine insensitive enzyme was inserted into the pigeonpea genome by Thu et al. (2007) through particle bombardment and *Agrobacterium* mediated transformation. They examined 11 lines which showed two- to sixfold increase in DHDPS activity compared to wild type in immature seeds at a late stage of development. In comparison to control lines, the *dhdps-r1* overexpression increased the free lysine concentration in pigeonpea seeds by 1.6–8.5 times.

Proline is an important amino acid in plants functions as an osmoprotectant and is crucial for maintaining osmotic balance, safeguarding enzymes and subcellular structures, and raising cellular osmolarity, which provides the turgor required for cell expansion under stressful circumstances. The rate-limiting enzyme in the production of proline, *1-pyrroline-5-carboxylate synthetase* (*P5CS*), is also inhibited by proline through feedback inhibition. Surekha et al. (2014) inserted a mutated version of *P5CS* named *P5CSF129A* from *Vigna aconitifolia* into pigeonpea genome. This mutated *P5CSF129A* gene is indifferent of feedback control. T_0 transgenic generation showed higher proline accumulation than control plants. A significant improvement was seen in chlorophyll content and growth performance in T_1 lines alongside decreased levels of lipid peroxidation. The relative water content under high salinity also showed improvement. *Render pest virus* (*RPVH*) and *peste des petits ruminants' virus* (*PPRV-HN*) both are the causal agents of devastating diseases in cattle animals with very high mortality rate such as cattle plague and Peste des Petits Ruminants respectively. New vaccination methods were developed using pigeonpea transformation to strengthen the immune systems of sheep, goats, and bovids against those viruses as the existing live attenuated vaccines are heat labile. Satyavathi et al. (2003) developed pigeonpea line that express *Rinderpest virus's hemagglutinin* protein. T_1 Pigeonpea leaves had the highest expression of the *hemagglutinin* protein at 0.49% of the total soluble protein. The transgene was expressed in the offspring of the fertile transgenic plants. Prasad et al. (2004) successfully generated transgenic pigeonpea lines by inserting two *PPRV* surface glycoproteins, hemagglutinin-neuraminidase, and fusion protein using pBI121 binary vector. T_1 plants showed transgene's inheritance.

Extracellular enzymes, especially those that cause the proteolytic breakdown of proteins in host plants are secreted by many phytopathogenic bacteria and some insects and crucial for pathogenesis. Plants have many inhibitors that work against these proteolytic enzymes as a key line of defense against these diseases. One such inhibitor named *cowpea protease inhibitor* (*CPI*), isolated from cowpea was inserted into pigeonpea genome through *Agrobacterium* mediated transformation. Transgenic pigeonpea lines showed higher level of defense against the lepidopteran insects (Lawrence and Koundal 2001).

11.2 Biofortification Resources of Pigeonpea Used in Other Transgenic Crops

In pigeonpea, under biotic and abiotic stress conditions, complex signaling pathways were found to be activated, causing changes in gene expression, necessary for plants

to adapt and acclimate. One such gene named *Pigeonpea hybrid-proline-rich protein encoding gene* (*CcHyPRP*) was used to develop transgenic tolerance lines in rice by Mellacheruvu et al. (2016). *CcHyPRP* was cloned under an inducible *rd29A* promoter and a constitutive *CaMV35S* promoter. Four independent homozygous T4 lines for each rd29ACcHyPRP and CaMV35SCcHyPRP were developed, which revealed very high accumulation of proline and endochitinase. In comparison to the control lines, the CcHyPRP transgenics showed greater resistance to rice blast disease causing fungus *Magnaporthe grisea*. Transgenic rice was shown to have more bZIP and endochitinase transcripts and endochitinase activity than control plants. These T_4 lines also demonstrated excellent levels of tolerance to the main abiotic stimuli, including heat, salinity, and drought, as demonstrated by enhanced chlorophyll content, survival rate, biomass, root, and shoot growth, in comparison to the untransformed lines. Additionally, under various biotic and abiotic stress situations, transgenic rice lines had larger panicles and more grains in comparison. In comparison to the control, the *CcHyPRP* transgenics showed increased catalase and superoxide dismutase (SOD) enzyme activity as well as decreased malondialdehyde (MDA) levels.

12 Future Prospects

In the post-green revolution period, improving the nutritional value of pigeonpea has become crucial for reducing malnutrition issues in developing nations. Establishing desired genotypes will be aided by in-depth knowledge of the genes and QTLs related to nutritional quality and seed quality (Singh et al. 2020). In order to develop molecular techniques aiming at enhancing seed quality and other nutritionally related qualities in pigeonpea, it will be essential to identify the genes/QTLs controlling the quality traits. To define quality features, attention should be paid to locate genetically varied and nutritionally improved pigeonpea lines (Singh et al. 2020). In order to measure various phenotypic features, it is crucial to design a high-throughput phenotyping platform. Examples of techniques that will be impactful for high throughput phenotyping include picture-based computer vision phenotyping, image processing, and data extraction tools. All integrated approaches will improve the understanding of systems biology by providing information on gene function, genomic architecture, organization, biological pathways, and metabolic and regulatory networks (Fig. 4).

The world's problems with malnutrition can be addressed in a new way by utilizing and combining cutting-edge NGS "omics" technology to sequence vast populations, uncover the genetic basis of agronomically essential traits, and anticipate breeding value. Breeders will be aided to gather information on specific alleles of known genes involved in nutritional grain quality attributes to achieve this goal through the availability of gene-based markers and cutting-edge techniques. Genomic regions/genes can be found that are expected to influence seed quality and nutritional qualities of interest by genotyping and phenotyping for those traits utilizing associations and machine learning models, drawing on the collection and

use of numerous unrelated lines. When omics technologies are used in conjunction with breeding programmes, it is anticipated that the nutritional quality of pigeonpea will improve.

References

Ade-Omowaye BIO, Tucker GA, Smetanska I (2015) Nutritional potential of nine underexploited legumes in Southwest Nigeria. Int Food Res J 22(2):798

Agus S, Achmadi SS, Mubarik NR (2017) Antibacterial activity of naringenin-rich fraction of pigeon pea leaves toward *Salmonella thypi*. Asian Pac J Trop Biomed 7(8):725–728

Ajaiyeoba EO, Ogbole OO, Abiodun OO, Ashidi JS, Houghton PJ, Wright CW (2013) Cajachalcone: an antimalarial compound from *Cajanus cajan* leaf extract. J Parasitol Res 2013:703781

Akinloye OA, Solanke OO (2011) Evaluation of hypolipidemic and potential antioxidant effects of pigeon pea (*Cajanus cajan* (L) Millsp.) leaves in alloxan-induced hyperglycemic rats. J Med Plant Res 5(12):2521–2524

Apata D (2008) Effect of cooking methods on available and unavailable carbohydrates of some tropical grain legumes. Afr J Biotechnol 7:2940–2945

Arora S, Mahato AK, Singh S, Mandal P, Bhutani S, Dutta S, Kumawat G, Singh BP, Chaudhary AK, Yadav R (2017) A high-density intraspecific SNP linkage map of pigeonpea (*Cajanas cajan* L. Millsp.). PLoS One 12:e0179747

Banu SA, Huda KMK, Tuteja N (2014) Isolation and functional characterisation of the promoter of a DEAD-box helicase Psp68 using *Agrobacterium*-mediated transient assay. Plant Signal Behav 9(6):e28992

Bohra A, Bharadwaj C, Radhakrishnan T, Singh NP, Varshney RK (2019) Translational genomics and molecular breeding for enhancing precision and efficiency in crop improvement programs: some examples in legumes. Indian J Genet 79:227–240

Brito SA, Rodrigues FF, Campos AR, da Costa JG (2012) Evaluation of the antifungal activity and modulation between *Cajanus cajan* (L.) Millsp. leaves and roots ethanolic extracts and conventional antifungals. Pharmacogn Mag 8(30):103–106

Dahiya JS (1987) Reversed-phase high-performance liquid chromatography of Cajanus cajan phytoalexins. J Chromatogr A 409(C):355–359

Dahiya JS (1991) Cajaflavanone and cajanone released from Cajanus cajan (L. Millsp.) roots induce nod genes of Bradyrhizobium sp. Plant Soil 134(2):297–304

Duker-Eshun G, Jaroszewski JW, Asomaning WA, Oppong-Boachie F, Brøgger Christensen S (2004) Antiplasmodial constituents of *Cajanus cajan*. Phytother Res 18(2):128–130

Ezike AC, Akah PA, Okoli CC, Okpala CB (2010) Experimental evidence for the antidiabetic activity of *Cajanus cajan* leaves in rats. J Basic Clin Pharm 1(2):81

Fanou BA, Klotoe JR, Fah L, Dougnon V, Koudokpon CH, Toko G, Loko F (2020) Ethnobotanical survey on plants used in the treatment of candidiasis in traditional markets of southern Benin. BMC Complement Med Ther 20(1):288

Fu YJ, Zu YG, Liu W, Hou CL, Chen LY, Li SM, Shi XG, Tong MH (2007) Preparative separation of vitexin and isovitexin from pigeonpea extracts with macroporous resins. J Chromatogr A 1139(2):206–213

Fu YJ, Liu W, Zu YG, Tong MH, Li SM, Yan MM, Efferth T, Luo H (2008) Enzyme assisted extraction of luteolin and apigenin from pigeonpea [Cajanus cajan (L.) Millsp.] leaves. Food Chemistry 111(2):508–512.

Fu Y, Kadioglu O, Wiench B, Wei Z, Gao C, Luo M, Gu C, Zu Y, Efferth T (2015) Cell cycle arrest and induction of apoptosis by cajanin stilbene acid from *Cajanus cajan* in breast cancer cells. Phytomedicine 22(4):462–468

Ganguly S, Ghosh G, Purohit A, Kundu Chaudhuri R, Chakraborti D (2018) Development of transgenic pigeonpea using high throughput plumular meristem transformation method. Plant Cell Tissue Organ Cult 135(1):73–83

Ghosh G, Purohit A, Chaudhuri RK, Chakraborti D (2014a) Advances in genetic transformation of important pulse crop pigeonpea. OA Biotechnol 12:5

Ghosh G, Purohit A, Ganguly S, Chaudhuri RK, Chakraborti D (2014b) In vitro shoot grafting on rootstock: an effective tool for *Agrobacterium*-mediated transformation of pigeonpea (*Cajanus cajan* (L.) Millsp.). Plant Biotechnol 31:301–308

Ghosh G, Ganguly S, Purohit A, Chaudhuri RK, Das S, Chakraborti D (2017) Transgenic pigeonpea events expressing Cry1Ac and Cry2Aa exhibit resistance to *Helicoverpa armigera*. Plant Cell Rep 36(7):1037–1051

Green PWC, Stevenson PC, Simmonds MSJ, Sharma HC (2003) Phenolic compounds on the pod-surface of pigeonpea, Cajanus cajan, mediate feeding behavior of Helicoverpa armigera larvae. J Chem Ecol 29(4):811–821

Gumisiriza H, Birungi G, Olet EA, Sesaazi CD (2019) Medicinal plant species used by local communities around Queen Elizabeth National Park, Maramagambo Central Forest Reserve and Ihimbo Central Forest Reserve, South western Uganda. J Ethnopharmacol 239:111926

Hassan EM, Matloub AA, Aboutabl ME, Ibrahim NA, Mohamed SM (2016) Assessment of anti-inflammatory, antinociceptive, immunomodulatory, and antioxidant activities of Cajanus cajan L. seeds cultivated in Egypt and its phytochemical composition. Pharmaceutical biology 54(8):1380–1391

Huang X, Han B (2014) Natural variations and genome-wide association studies in crop plants. Annu Rev Plant Biol 65:531–551

Hussain AI, Pavithra HV, Sreevathsa R, Nataraja KN, Babu N (2016) Development of transgenic pigeonpea (*Cajanus cajan* L. Millsp.) overexpressing citrate synthase gene for high phosphorus uptake. Indian J Exp Biol 54(8):493–501

Iweala EE, Evbakhavbokun WO, Maduagwu EN (2019) Antioxidant and hepatoprotective effect of *Cajanus cajan* in *N*-nitrosodiethylamine-induced liver damage. Sci Pharm 87(3):24

Khan R, Soni LK, Jain S (2015) Anthelmintic activity of leaves of *Cajanus cajan* Linn on Indian earthworm. Asian J Pharm Health Sci 5(4):1327–1330

Koné WM, Koffi AG, Bomisso EL, Tra Bi FH (2011) Ethnomedical study and iron content of some medicinal herbs used in traditional medicine in Cote d'Ivoire for the treatment of anaemia. Afr J Tradit Complement Altern Med 9(1):81–87

Kumar V, Khan AW, Saxena RK, Garg V, Varshney RK (2016) First-generation HapMap in *Cajanus* spp. reveals untapped variations in parental lines of mapping populations. Plant Biotechnol J 14(8):1673–1681

Lans C (2007) Comparison of plants used for skin and stomach problems in Trinidad and Tobago with Asian ethnomedicine. J Ethnobiol Ethnomed 3:3

Lawrence PK, Koundal KR (2001) *Agrobacterium tumefaciens* mediated transformation of pigeonpea (*Cajanus cajan* L. Millsp.) and molecular analysis of regenerated plants. Curr Sci 80:1428–1432

Liu HJ, Yan J (2019) Crop genome-wide association study: a harvest of biological relevance. Plant J 97:8–18

Luo QF, Sun L, Si JY, Chen DH (2008) Hypocholesterolemic effect of stilbenes containing extract-fraction from *Cajanus cajan* L. on diet-induced hypercholesterolemia in mice. Phytomedicine 15(11):932–939

Luo M, Liu X, Zu Y, Fu Y, Zhang S, Yao L, Efferth T (2010) Cajanol, a novel anticancer agent from pigeonpea [*Cajanus cajan* (L.) Millsp.] roots, induces apoptosis in human breast cancer cells through a ROS-mediated mitochondrial pathway. Chem Biol Interact 188(1):151–160

Mellacheruvu S, Tamirisa S, Vudem DR, Khareedu VR (2016) Pigeonpea hybrid-proline-rich protein (CcHyPRP) confers biotic and abiotic stress tolerance in transgenic rice. Front Plant Sci 6:1167

Mir RR, Kudapa H, Srikanth S, Saxena RK, Sharma A, Azam S, Saxena K, Varma Penmetsa R, Varshney RK (2014) Candidate gene analysis for determinacy in pigeonpea (*Cajanus* spp.). Theor Appl Genet 127:2663–2678

Nahar L, Nasrin F, Zahan R, Haque A, Haque E, Mosaddik A (2014) Comparative study of antidiabetic activity of *Cajanus cajan* and *Tamarindus indica* in alloxan-induced diabetic mice with a reference to in vitro antioxidant activity. Pharmacognosy Res 6(2):180–187

Nandini B, Reddy UG, Mallikarjuna BP, Manu B, Vaijayanthi PV, Ashwini M, Chakraborti D (2022) Genomic design for abiotic stress resistance in pigeonpea. In: Kole C (ed) Genomic design for abiotic stress resistant pulse crops. Springer, Cham, pp 169–248

Negi J, Rathinam M, Sreevathsa R, Kumar PA (2021) Transgenic pigeonpea [*Cajanus cajan* (L). Millsp.]. In: Kavi Kishor PB, Rajam MV, Pullaiah T (eds) Genetically modified crops: current status, prospects and challenges, vol 1. Springer, Singapore, pp 79–96

Nicholson RA, David LS, Pan RL, Liu XM (2010) Pinostrobin from *Cajanus cajan* (L.) Millsp. inhibits sodium channel-activated sdepolarisation of mouse brain synaptoneurosomes. Fitoterapia 81(7):826–829

Nix A, Paull CA, Colgrave M (2015) The flavonoid profile of pigeonpea, Cajanus cajan: a review. SpringerPlus 4:1–6

Nwodo UU, Ngene AA, Iroegbu CU, Onyedikachi OA, Chigor VN, Okoh AI (2011) In vivo evaluation of the antiviral activity of *Cajanus cajan* on measles virus. Arch Virol 156(9):1551–1557

Obala J, Saxena RK, Singh VK, Kumar CV, Saxena KB, Tongoona P, Sibiya J, Varshney RK (2019) Development of sequence-based markers for seed protein content in pigeonpea. Mol Genet Genomics 294:57–68

Okigbo RN, Omodamiro OD (2007) Antimicrobial effect of leaf extracts of pigeon pea (*Cajanus cajan* (L.) Millsp.) on some human pathogens. J Herbs Spices Med Plants 12(1-2):117–127

Olorunnisola OS, Adetutu A, Balogun EA, Afolayan AJ (2013) Ethnobotanical survey of medicinal plants used in the treatment of malaria in Ogbomoso, Southwest Nigeria. J Ethnopharmacol 150 (1):71–78

Pahwa K, Ghai N, Bedi S (2013) Pigeonpea: a potential multipurpose crop. Rastriya Krishi 8:3

Pal D, Mishra P, Sachan N, Ghosh AK (2011) Biological activities and medicinal properties of *Cajanus cajan* (L.) Millsp. J Adv Pharm Technol Res 2(4):207

Patel NK, Bhutani KK (2014) Pinostrobin and *Cajanus* lactone isolated from *Cajanus cajan* (L.) leaves inhibits TNF-α and IL-1β production: in vitro and in vivo experimentation. Phytomedicine 21(7):946–953

Patil PG, Dubey J, Bohra A, Mishra RK, Saabale PR, Das A, Rathore M, Singh NP (2017) Association mapping to discover significant marker-trait associations for resistance against Fusarium wilt variant 2 in pigeonpea [*Cajanus cajan* (L.) Millspaugh] using SSR marker. J Appl Genet 58:307–319

Pazhamala L, Saxena RK, Singh VK, Sameerkumar CV, Kumar V, Sinha P, Patel K, Obala J, Kaoneka SR, Tongoona P, Shimelis HA (2015) Genomics-assisted breeding for boosting crop improvement in pigeonpea (*Cajanus cajan*). Front Plant Sci 17(6):50

Pfeiffer WH, McClafferty B (2007) HarvestPlus: breeding crops for better nutrition. Crop Sci 47:S-88–S-105

Prasad V, Satyavathi VV, Sanjaya, Valli KM, Khandelwal A, Shaila MS, Lakshmi Sita G (2004) Expression of biologically active hemagglutinin-neuraminidase protein of Peste des petits ruminants virus in transgenic pigeonpea [*Cajanus cajan* L. Millsp.]. Plant Sci 166:199–205

Rahmatullah M, Mukti IJ, Haque AK, Mollik MA, Parvin K, Jahan R, Chowdhury MH, Rahman T (2009) An ethnobotanical survey and pharmacological evaluation of medicinal plants used by the Garo tribal community living in Netrakona district, Bangladesh. Adv Nat Appl Sci 3(3):402–418

Reddy LJ, Green JM, Singh U, Bisen SS, Jambunathan R (1979) Seed protein studies on *Cajanus cajan*, *Atylosia* spp. and some hybrid derivatives. In: Proceedings of the international symposium on seed protein improvement in cereals and grain legumes, vol II. International Atomic Energy Agency, Vienna, pp 105–117

Salehi B, Ata AV, Anil Kumar N, Sharopov F, Ramirez-Alarcon K, Ruiz-Ortega A, Sharifi-Rad J (2019) Antidiabetic potential of medicinal plants and their active components. Biomolecules 9 (10):551

Satyavathi VV, Prasad V, Khandelwal A, Shaila MS, Lakshmi Sita G (2003) Expression of hemagglutinin protein of Rinderpest virus in transgenic pigeon pea (*Cajanus cajan* (L.) Millsp.) plants. Plant Cell Rep 21:651–658

Saxena KB (2008) Genetic improvement of pigeon pea – a review. Trop Plant Biol 1(2):159–178

Saxena KB, Kumar RV (2003) Development of a cytoplasmic nuclear male-sterility system in pigeonpea using *C. scarabaeoides* (L.) Thouars. Indian J Genet Plant Breed 63:225–229

Saxena KB, Sawargaonkar SL (2015) Genetic enhancement of seed proteins in pigeonpea – methodologies, accomplishments, and opportunities. Int J Sci Res 4(5):03–07

Saxena KB, Sharma D (1990) Pigeonpea: genetics. In: Nene YL, Hall SD, Sheila VK (eds) The pigeonpea. CAB International, Wallingford, pp 137–157

Saxena KB, Kumar RV, Rao PV (2002) Pigeonpea nutrition and its improvement. J Crop Prod 5(1/2):227–260

Saxena RK, Varma Penmetsa R, Upadhyaya HD, Kumar A, Carrasquilla-Garcia N, Schlueter JA, Farmer A, Whaley AM, Sarma BK, May GD (2012) Large-scale development of cost-effective single-nucleotide polymorphism marker assays for genetic mapping in pigeonpea and comparative mapping in legumes. DNA Res 19:449–461

Saxena RK, Von Wettberg E, Upadhyaya HD, Sanchez V, Songok S, Saxena K, Kimurto P, Varshney RK (2014) Genetic diversity and demographic history of *Cajanus* spp. illustrated from genome-wide SNPs. PLoS One 9:e88568

Saxena KB, Tikle AN, Kumar RV, Choudhary AK, Bahadur B (2016) Nectarivore-aided shybridisation and its exploitation for productivity enhancement in pigeonpea. Int J Sci Res Publ 6(08):321–331

Saxena RK, Kale SM, Kumar V, Parupali S, Joshi S, Singh V, Garg V, Das RR, Sharma M, Yamini KN (2017) Genotyping-by-sequencing of three mapping populations for identification of candidate genomic regions for resistance to sterility mosaic disease in pigeonpea. Sci Rep 7: 1–10

Saxena RK, Rathore A, Bohra A, Yadav P, Das RR, Khan AW, Singh VK, Chitikineni A, Singh IP, Kumar CS (2018) Development and application of high-density Axiom *Cajanus* SNP array with 56K SNPs to understand the genome architecture of released cultivars and founder genotypes. Plant Genome 11:180005

Saxena RK, Saxena KB, Varshney RK (2019) Pigeonpea (*Cajanus cajan* L. Millsp.): an ideal crop for sustainable agriculture. In: Al-Khayri J, Jain S, Johnson D (eds) Advances in plant breeding strategies: legumes. Springer, Cham, pp 409–429

Shen J, Li H, Neumann G, Zhang F (2005) Nutrient uptake, cluster root formation and exudation of protons and citrate in *Lupinus albus* as affected by localised supply of phosphorus in a split-root system. Plant Sci 168(3):837–845

Singh U, Eggum BO (1984) Factors affecting the protein quality of pigeonpea (*Cajanus cajan* L.). Plant Food Hum Nutr 34(4):273–283

Singh N, Rai V, Singh NK (2020) Multi-omics strategies and prospects to enhance seed quality and nutritional traits in pigeonpea. Nucleus 63:249–256

Surekha CH, Kumari KN, Aruna LV, Suneetha G, Arundhati A, Kavi Kishor PB (2014) Expression of the *Vigna aconitifolia* P5CSF129A gene in transgenic pigeonpea enhances proline accumulation and salt tolerance. Plant Cell Tissue Organ Cult 116(1):27–36

Teixeira EM, Silva-López RE, Da Silva BR, Fontão AP, Sampaio AL (2021) *Cajanus cajan* (L.) Millsp. aqueous extracts against melanoma cell line and their proteases. Eur J Med Plants 32(2): 1–14

Thu TT, Dewaele E, Trung LQ, Claeys M, Jacobs M, Angenon G (2007) Increasing lysine levels in pigeonpea (*Cajanus cajan* (L.) Millsp.) seeds through genetic engineering. Plant Cell Tissue Organ Cult 91(2):35–143

Toklu F, Sen Gupta D, Karaköy T, Özkan H (2021) Bioactives and nutraceuticals in food legumes: nutritional perspective. In: Breeding for enhanced nutrition and bio-active compounds in food legumes. Springer, Cham, pp 229–245

Vales MI, Srivastava RK, Sultana R, Singh S, Singh I, Singh G, Saxena KB (2012) Breeding for earliness in pigeonpea: development of new determinate and nondeterminate lines. Crop Sci 52 (6):2507–2516

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MT, Azam S, Fan G, Whaley AM (2012) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotechnol 30:83

Varshney RK, Mohan SM, Gaur PM, Gangarao N, Pandey MK, Bohra A, Sawargaonkar SL, Chitikineni A, Kimurto PK, Janila P (2013) Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. Biotechnol Adv 31:1120–1134

Varshney RK, Saxena RK, Upadhyaya HD, Khan AW, Yu Y, Kim C, Rathore A, Kim D, Kim J, An S (2017) Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. Nat Genet 49:1082–1088

Vo TT, Yang NC, Yang SE, Chen CL, Wu CH, Song TY (2020) Effects of *Cajanus cajan* (L.) Millsp. roots extracts on the antioxidant and anti-inflammatory activities. Chin J Physiol 63(3): 137–148

Wang D, Xiao M, Li Y, Shen X, Lu Y, Liu K, Li Z, Hu Y (2011) Determination of longistylin A and longistylin C in *Cajanus cajan*. Zhongguo Zhong Yao Za Zhi 36(19):2680–2683

Wei ZF, Luo M, Zhao CJ, Li CY, Gu CB, Wang W, Zu YG, Efferth T, Fu YJ (2013) UV-induced changes of active components and antioxidant activity in postharvest pigeon pea [Cajanus cajan (L.) Millsp.] leaves. J Agric Food Chem 61(6):1165–1171

Wu J, Li B, Xiao W, Hu J, Xie J, Yuan J, Wang L (2020) Longistylin A, a natural stilbene isolated from the leaves of *Cajanus cajan*, exhibits significant anti-MRSA activity. Int J Antimicrob Agents 55(1):105821. (Erratum in: Int J Antimicrob Agents 58(2):106392, 2021)

Yadav P, Saxena KB, Hingane A, Kumar CV, Kandalkar VS, Varshney RK, Saxena RK (2019) An "Axiom *Cajanus* SNP array" based high density genetic map and QTL mapping for high-selfing flower and seed quality traits in pigeonpea. BMC Genomics 20:1–10

Zu YG, Fu YJ, Liu W, Hou CL, Kong Y (2006) Simultaneous determination of four flavonoids in pigeonpea Cajanus cajan (L.) Millsp. leaves using RP-LC-DAD. Chromatographia 63(9–10):499–505

Nutrigenomics of Mungbean

B. Manu, Jayashree Ugalat, P. R. Saabale, Revanappa Biradar, Suma C. Mogali, and Shivanand Koti

Contents

B. Manu (✉) · P. R. Saabale · R. Biradar
ICAR-Indian Institute of Pulses Research, Regional Station, Dharwad, Karnataka, India
e-mail: Manu.B@icar.gov.in

J. Ugalat
Department of Biotechnology and Crop Improvement College of Horticulture, UHSB Campus, Bangalore, India

S. C. Mogali
AICRP-MULLaRP Scheme, University of Agricultural Sciences, Dharwad, Karnataka, India

S. Koti
Department of Fruit Science College of Horticulture, UHSB Campus, Bangalore, India

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_29

Abstract

Mungbean (*Vigna radiata* (L.) Wilczek), one of the important leguminous crops, contains balanced nutrients protein, dietary fiber, minerals, vitamins, and good amounts of bioactive compounds such as phenolic acids, flavonoids, and tannins. The discipline of nutriomics or nutrigenomics explains the reciprocal interaction of nutrients and genes at the molecular level. Nutrigenomics can be helpful for better perceptiveness of nutrient-gene interactions and the development of "personalized nutrition" for good health and disease prevention. Considering the worldwide importance of mungbean, there is a scope for improving its nutritional value. Crop improvement focuses on enhancing protein and starch content and quality, the content of minerals like iron, zinc, and also in removing the anti-nutritional compounds like phytic acid. The mungbean whole-genome sequence data helps in the advancement of genomics research in Vigna species and speed up the mungbean breeding programs. Genotyping has enabled marker-assisted selection (MAS) and identified SNPs, serve as important resources to facilitate MAS for nutritional improvement. 8Sα globulin or vicilin is the major storage protein of mungbean and was engineered using site-directed mutagenesis. Transcriptomic, metabolomic, and ionomics analyses reveal the molecular mechanisms of nutrient accumulation, nutritional/nutraceutical value and health-promoting properties, and bioavailable minerals.

Keywords

Nutrition · Nutrigenomics · Transcriptomics · Metabolomics · Ionomics · Disease · Health

1 Introduction

Mungbean (*Vigna radiata* (L.) Wilczek) is a low-input, short-season, important grain legume of tropical countries and is also known as greengram and moong. This crop is grown in various climatic conditions, locations, and seasons due to its inherent intrinsic tolerance mechanisms. It is grown over seven million hectares in South and Southeast Asia, Australia, Africa, and South America. It is cultivated as an intercrop or subsistence monocrop predominantly by small and marginal farmers; it is consumed as whole seed or split cooking, flour, or as sprouts; and it is a cheap source of vegetable protein, minerals, vitamins, and dietary fiber for the vegetarian population. The crop has valuable nutritional and health benefits, where malnutrition is a key concern. It has the ability to improve the soil health by fixing atmospheric nitrogen through a symbiotic association with *Rhizobium*, thus enhancing the yield of the subsequent crops. Other major benefits of this crop to farmers are its high nutritional and monetary value, as it fits into intensive wheat, rice, and summer cereal rotation systems, short duration, photo-thermo insensitivity, low input requirement, and heat and drought stress tolerance. Development of extra-early and yellow mosaic-resistant varieties paved the way for extensive cultivation

of the crop in several parts of India into different cropping systems (rice-fallow, rice-rice, rice-wheat, and rice-maize) and for cultivation in different parts of the world including South America and sub-Saharan African regions (Moghadam et al. 2011). The yield potential of mungbean is 2.5–3.0 t/ha; however, the global productivity of mungbean is only 0.5 tons per hectare, impacted due to traditional low input farming system, nonavailability of quality seeds of improved varieties to farmers, and biotic and abiotic stress factors. The major biotic factors include insect pests, diseases, and weeds (Pandey et al. 2018; Nair et al. 2019). Insect pests cause direct damage by feeding the crop and indirect damage by transmitting viral diseases. Abiotic stresses, waterlogging, salinity, heat, and drought stress are key concerns in mungbean production (Hanumantha Rao et al. 2016).

During the eighteenth century, the "analytical chemistry era" Lavoisier discovered how carbon dioxide, water, and energy were generated after food digestion in the body (Vasconcelo 2010). In the "biological era" around the nineteenth century, where important discoveries in metabolism and chemistry were done, helping the science of nutrition to prevent metabolic disorders (Cruz et al. 2003). Dietary components/nutritional attributes are environmental factors that can interact with the genome which determines the health condition of individuals (Ronteltap et al. 2008). With the advancement of science, the discipline of nutriomics or nutrigenomics was introduced to get insights regarding how could food bioactive molecules and genes influence the health of an individual positively and negatively (C. Kole and AG Abbott 2011 for ICPN at PAG 2011; Dauncey 2014; Cozzolino and Cominetti 2013). Nutrigenomics is an area of nutrition that corresponds to the use of physiology, biochemistry, genomics, metabolomics, proteomics, transcriptomics, nutrition, and epigenomics to explain the reciprocal interaction of nutrients and genes at the molecular level (Dauncey 2014; Cozzolino and Cominetti 2013). The study of nutrigenomics has progressed considerably in recent times, nutrigenomics aims at understanding the impact of nutritional factors in protecting the genome. Studies on folate metabolism proved that folic acid is a nutrient, ensures the genetic integrity of an individual, and acts as a cofactor for the biosynthesis of nucleotides and thymidylate (Liu and Qian 2011; Cozzolino and Cominetti 2013). Bioactive molecules contained in the diet act as cofactors/substrates in DNA metabolism and repair. Studies showed that vitamins A, D, and fatty acids activate nuclear receptors by direct action and induce gene transcription, and some of the bioactive compounds influence molecular signaling pathways (Fialho et al. 2008; Cozzolino and Cominetti 2013).

2 Nutritional Profile of Mungbean

Mungbean (*Vigna radiata* L.), an important pulse crop consumed across the world, especially in Asian continents, has a long history of usage in traditional medicine. Mungbean is consumed as a food for centuries. Mungbean contains balanced nutrients, importantly protein, minerals, dietary fiber, vitamins, and enough amounts of bioactive compounds (Gan et al. 2017). For those individuals who are vegan or cannot

afford animal proteins, the mungbean is a relatively low-cost and has a good amount of protein. Furthermore, this protein is easily digestible in comparison protein of other legumes (Mubarak 2005; Yi-Shen et al. 2018), that is the how, it is becoming as a popular functional food in promoting good health (Tables 1, 2, and 3).

The mungbean is a rich source of polyphenolics. The important phenolic contents in the mungbean are phenolic acids, tannins, and flavonoids (Lee et al. 2011; Shi et al. 2016; Singh et al. 2017a). The seed coats and cotyledons of mungbeans contain phenolics; however, most are present in the coats of seed. The most abundant secondary metabolite in the mungbean is flavonoids. The five subclasses of

Table 1 Macronutrient composition of mungbean (Dahiya et al. 2015a)

Macronutrient	Average[a]	Minimum	Maximum
Moisture (g/100 g)	9.80	4.10	15.20
Crude protein content (g/100 g dm)	23.8	14.6	32.6
Crude fat (g/100 g dm)	1.22	0.71	1.85
Crude fiber (g/100 gdm)	4.57	3.8	6.15
Ash (g/100 gdm)	3.51	0.17	5.87
Total carbohydrate (g/100 g dm)	61.0	53.3	67.1
Energy yield (kcal/100 g dm)	344	338	347

[a]Average of the values quoted by several authors

Table 2 Content of amino acids in mungbean protein isolates (Yi-Shen et al. 2018)

Amino acids	MBPI (mgg^{-1})
Total amino acids content	800.2
Total essential amino acids content	348.2 (43.51%)[a]
Total aromatic amino acids content	96.7 (12.08%)[a]
Total sulfur containing amino acids	13.0 (1.63%)[a]
Phenylalanine+tyrosine	90.3
Leucine	74
Lysine	62.4
Valine	46.3
Isoleucine	39.1
Histidine	27.9
Threonine	28.4
Methionine+cysteine	13
Tryptophan	6.4
Glutamic acid/glutamine	125.4
Aspartic acid/asparagine	85.3
Arginine	64.4
Serine	38.5
Alanine	36.6
Glycine	32.2
Proline	30

MBPI mungbean protein isolates
[a]Percent amino acids relative to total amino acids in MBPI

Table 3 Mungbean mineral nutrients content (Habibullah et al. 2007)

Mineral	Content (mg/100 g)
Na	22 mg/100 g
K	1443 mg/100 g
Ca	216 mg/100 g
Mg	204 mg/100 g
P	374 mg/100 g
Fe	11.34 mg/100 g
Zn	1.88 mg/100 g
Cu	1.92 mg/100 g
Mn	1.49 mg/100 g
Pb	2.64 mg/100 g

flavonoids in the mungbean are, i.e., flavonols, flavones, isoflavonoids, flavanols, and anthocyanins. Flavones (isovitexin, vitexin, isovitexin-6″-*O*-α-l-glucoside, and luteolin) and flavonols (myricetin, quercetin, and kaempferol) are the most occurring flavonoids found in the mungbean. Isovitexin and vitexin were considered to be the two major flavonoids in the seed of mungbean; their contents in the seed coat are attributed to 95.6% and 96.8% of the total vitexin and isovitexin, correspondingly (Cao et al. 2011; Peng et al. 2008). Quantification by chemical analysis indicated that the contents of isovitexin and vitexin in the seed coat were high, 37.43 mg/g and 47.18 mg/g, correspondingly.

Legumes constitute the most abundant and least expensive protein source for human and animal diets; however, their utilization is limited largely because of the anti-nutritional/anti-physiological compounds (Liener 1994) present in these. Those are protease inhibitors (trypsin), amylase inhibitors, lectins, polyphenols, tannins (TN), phytic acid (PA), flatulence factors, and allergens (Liener and Kakade 1980; Liener 1995). These factors downgrade the nutritive value of beans through direct and indirect reactions, inhibit protein and carbohydrate digestibility, induce pathological conditions in the intestine and liver tissue, and finally affect metabolism, inhibit enzymes, and bind to nutrients, making them unavailable (Bressani et al. 1989; Bressani 1993). Anti-nutritional factors: trypsin inhibitor (TIUA; trypsin inhibitor units /mg protein) is 15.8, hemagglutinin activity (HUB; hemagglutinin units/g) is 2670, tannins (mg/g) 3.30, and phytic acid (mg/gm) is 5.80.

3 Approaches of Biofortification Through Omics Methods

3.1 Nutritional Genomics and Epigenomics

Nutrigenomics is the study of the gene-nutrient interaction, and it shows that some nutrients, called bioactive compounds, can shape the genetic expression or change the nucleotide sequence. To analyze the interaction between genes and nutrients, the term "nutrigenomics" was proposed. Hence, nutrigenomics makes use of biochemistry, physiology, nutrition, genomics, proteomics, metabolomics, transcriptomics,

and epigenomics to get and explain the interactions between genes and nutrients at a chemical level. Nutrigenomics can be helpful for better perceptiveness of nutrient-gene connections and the development of "personalized nutrition" for good health and disease prevention (Di Renzo et al. 2019).

3.2 Important Traits and Breeding Goals

Mungbean is having a considerably good amount of protein, carbohydrates, and minerals such as zinc and Iron. Considering its worldwide importance, the crop is having a scope for improving its nutritional value with the available germplasm (Singh 2013). Studies must be conducted to understand the diversity at the level of nutrients in the mungbean germplasms. The increasing need for plant-based protein foods is a scope to study the functional properties of mungbean protein and starch (Shrestha et al. 2022). The mungbean breeding program should expand research to nutritional and food processing properties. Studies have to be conducted on several factors including the variety used, the region where the crop is grown, agronomic practices that have been adopted, and the storage environment. Further, studies have to be undertaken on the postharvest processes such as sprouting, dehulling, soaking, boiling, autoclaving, and microwave cooking to analyze their effect on the composition of nutritional and anti-nutritional contents of mungbean. Crop improvement in mungbean should further focus on enhancing protein and starch content and quality, the content of minerals like iron and zinc, and also in removing the anti-nutritional compounds like phytic acid (Samtiya et al. 2020). Micronutrients such as iron and zinc should be biofortified, for which identification of suitable parents, improvement of populations, and identification of quantitative trait loci (QTLs) for marker-assisted selection (MAS) are essential.

3.3 Genome Size and Genomic Resources

Mungbean is a member of the subgenus Ceratotropis of the genus Vigna, which also includes several agriculturally significant legumes, including rice bean, black gram, adzuki bean, and moth bean. While the majority of Vigna species are diploid (Egawa and Tomooka 1994), *Vigna reflexo-pilosa* ($2n = 4x = 44$) is a tetraploid species. The genomic sizes of the Vigna species range from 416 to 1394 Mb (Parida et al. 1990; Lakhanpaul and Babu 1991).

A consortium of 12 universities, led by Suk-Ha-Lee of Seoul National University in Korea, completed the draft genome sequencing of the cultivated mungbean (*V. radiata var. radiata* VC1973A), using two next-generation sequencing platforms HiSeq2000 and GS FLX+ were employed to sequence and assemble the information into 11 pseudo chromosomes. For a thorough understanding of genus Vigna's polyploidization, speciation, and domestication processes, whole-genome sequences of a wild relative mungbean (*V. radiata var. sublobata*) and a tetraploid relative of mungbean (*V. reflexo-pilosa* var. *glabra*), similarly transcriptome sequences of 22 Vigna accessions of 18 species were deciphered. The scientists also constructed a genetic map based on F6 population of 190 recombinant inbred lines (RIL) using genotyping by sequencing

Table 4 List of genomics databases/resources

S. No	Name of database	Source link/URL
1	Next Gen Seek	http://nextgenseek.com
2	Kevin's GATTACA World	http://kevin-gattaca.blogspot.com
3	In Between Lines of Code	http://flxlexblog.wordpress.com
4	Next-Gen Sequencing	http://nextgenseq.blogspot.com
5	CoreGenomics	http://core-genomics.blogspot.com
6	RNA-Seq Blog	http://www.rna-seqblog.com
7	Next Generation Technologist	http://www.yuzuki.org
8	Blog @ Illumina	http://blog.illumina.com
9	Bits of DNA	http://liorpachter.wordpress.com/seq
10	Journal of Next Generation Sequencing& Applications	http://www.omicsonline.org/nextgenerationsequencing-applications.php
11	Omics! Omics!	http://omicsomics.blogspot.com
12	PlantGDB	www.plantgdb.org/MtGDB/
13	LIS – Legume Information System	http://legumeinfo.org/gbrowsecajca1.0
14	Phytozome 10.2	http://phytozome.jgi.doe.gov/commonbean
15	Legume Information System	http://cicar.comparative-legumes.org/
16	Mungbean Genome Jbrowse	http://plantgenomics.snu.ac.kr/
17	Adzuki bean Genome Jbrowse	http://plantgenomics.snu.ac.kr/
18	PlantGDB	
19	Lotus japonicus genome assembly build 2.5	http://www.kazusa.or.jp/lotus/
20	PlantGDB	www.plantgdb.org/MtGDB/
21	Phytozome 10.2	http://phytozome.jgi.doe.gov/soybean

(GBS) strategy. The work of Kang et al. (2014) provided a deeper understanding of the evolutionary history of Vigna spp. (particularly subgenus Ceratotropis) based on comparative genomics strategy, with the 421 Mb (80%) assembled genome coupled with sequence information of related Vigna species. Vigna species may be utilized as model legume plants in genetic studies to shed light on crop domestication and species divergence due to their short life cycle and small genome size. The mungbean whole-genome sequence data will help in the advancement of genomics research in Vigna species and speed up mungbean breeding programs, which could serve as a model for future efforts to resequence the Vigna germplasm. Scientists have also developed webserver/repositories (Table 4) containing the genomics and related information of selected Vigna species, like Vigna Genome Server (Vig GS) which incorporates annotated exon-intron structures, along with evidence for transcripts and proteins, visualized in GBrowse (Sakai et al. 2016).

3.4 Molecular Mapping and Breeding

DNA-based marker systems have become available over the past three decades. These include restriction fragment length polymorphisms (RFLPs), SSRs or microsatellites,

random amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLPs), single nucleotide polymorphisms (SNPs), and diversity array technology (DArT). Among these markers, RAPD, AFLP, and RFLP are frequently utilized for marker-trait association and analysis of pulse diversity. However, plant breeders do not prefer their use for MAS because of the difficulty in handling, low reproducibility, and use of radioactive elements for generating these markers (Gupta et al. 2010). Only PCR (polymerase chain reaction)-based SSR and SNP are preferred by breeders, as these markers can easily be employed in genotyping of large populations. Also, more reproducibility and ease in usage make them preferential to conventional plant breeders for MAS (Gupta et al. 2010). These have been extensively utilized (Kumar et al. 2011) in many crops; however, their use is still inadequate in pulses (Varshney et al. 2009; Saxena et al. 2010). Therefore, greater focus is being put to develop more markers for pulses, which are considered orphan legumes (Hamwieh et al. 2009; Varshney et al. 2009). Close phylogenetic similarity has spurred researchers to transfer SSR markers from one pulse crop to another to lower the cost of developing these markers (Datta et al. 2010; Reddy et al. 2010.) Different types of markers such as RFLPs, RAPDs, AFLPs, SSRs, and ISSRs have been used in discerning genetic diversity and developing linkage maps in these crops. In mungbean, eight genetic linkage maps have been developed so far but no map contained enough markers to resolve all the 11 linkage groups.

Scientists are also working on genome-wide identification of SNP and association mapping of seed mineral concentration in mungbean (*Vigna radiata* L.). To assess the diversity of mungbeans available to breeders in the United States and to execute a genome-wide association study (GWAS), Wu et al. (2020) used genotyping by sequencing (GBS) tool. This study identified high-quality single nucleotide polymorphisms (SNPs) a total of 6486 from the GBS dataset and found marker x trait associations (MTAs) with calcium, iron, manganese, potassium phosphorous, zinc, or sulfur contents in mungbean grain. The 43 MTAs spread across 35 genomic regions elucidating an average of 22% of the variation for each seed nutrient. SNPs identified will serve as important resources of marker-assisted selection (MAS) for nutritional value in the mungbean.

4 Metabolomics

Plants produce wide numbers of nutrients that impart synergistic relations among the different combinations of nutrients. Therefore, inclusive nutrient profiling is required to evaluate the nutritional/nutraceutical value and health-promoting properties of crops. To acquire such datasets of mungbean, which is well known as a medicinal crop with heat-alleviating character, metabolites profiling is essential. Metabolomic and proteomic analysis of four genotypes from China, Thailand, and Myanmar has indicated a total of 449 proteins and 210 metabolic compounds in the seed coat. The first complete dataset of mungbean for nutraceutical values has indicated 480 proteins, and 217 metabolic factors in seed flesh. Whereas, gel-free/label-free proteomic analysis and metabolomic analysis in combination and pathway reconstruction

detected that amino acid metabolism is more predominant in flesh. Compared to flesh coat contains a wider variety of lipids and phenolic acids/flavonoids. Among the compounds detected in the coat, sphingolipids, arachidonic acid, and prostaglandin E2 are related to defense response induction. Furthermore, the identification of prostaglandin F2α added support to the empirical validity of the usage of mungbean. The abundance of bioactive compounds such as naringenin, which can be metabolized into vitexin, varied among genotypes. Lipids together with flavonoids may be possibly responsible compounds for the biological activity of mungbean coat and flesh (Hashiguchi et al. 2017).

5 Nutritional Transcriptomics

The recent availability of genome sequence information of several legume crops has led to boosting genomics research. Study of the transcriptome at a global level can provide insights into the gene space, gene function, transcriptional programs, and molecular basis of various cellular events, even in the absence of genome sequence (Garg and Jain 2013). Transcriptome analysis has been shown as an indispensable step for basic and applied research in any living system. High-quality sequence reads of cDNA obtained using sequencing technology, Illumina paired-end sequences, were assembled into unigenes (48,693) with a length of 874 bp, on an average. Among these unigenes, 25,820 (53.0%) and 23,235 (47.7%) indicated significant similarity to nonredundant protein and nucleotide sequence databases in the NCBI, respectively. A set of unigenes, 19,242 (39.5%), were classified into gene ontology categories, whereas 18,316 (37.6%) were classified into Swiss-Prot categories and 10,918 (22.4%) into KOG database categories (E-value $<$1.0E-5). By using the Kyoto Encyclopedia of Genes and Genome (KEGG) pathway database, a total of 6585 (8.3%) were mapped onto 244 pathways (Jain et al. 2020). During legume development, gene expression profiles were closely related to the ascorbic acid and phenolics accumulation regularity. The gene expression profiles of 25 key-coding genes in ascorbic acid and phenolics metabolic pathways, as well as the dynamitic changes of ascorbic acid, phenolic profiles, and antioxidant activities with the legume. From 8 to 17 days after flowering (DAF), *VrVTC2* and *VrGME* played important role in ascorbic acid accumulation. *VrPAL* and *VrCHS* were shown positive correlations with daidzein and glycitin accumulation, and *VrIFS* had a strong positive correlation in glycitin synthesis. Ascorbic acid and phenolics compounds dramatically increased the antioxidant properties during the mungbean maturation stage (Lu et al. 2019).

Molecular mechanisms, involved in the primary metabolic pathways and regulation of post-germination seedling growth, are understood in mungbean. In mungbean, during post-germination seedling growth, various metabolic and physiological changes occur, leading to the improvement of its nutritional values. Transcriptomic and metabolomic analyses of mungbean samples from 6-hour, 3-day, and 6-day after imbibition (6-HAI, 3-DAI, and 6-DAI) revealed the primary metabolites regulatory mechanism during the post-germination seedling growth.

Rapid changes in the transcript level of starch and sucrose metabolism were observed from 6-HAI to 3-DAI, glycolysis, citrate cycle, amino acids synthesis, and plant hormones regulation (Wang et al. 2020). An increase in transcript levels has led to later changes in the metabolites, including carbohydrates and amino acids. During this process, most amino acids and monosaccharides went on increasing and accumulated in germinated sprouts for 6 days. Changes in hormones including abscisic acid, gibberellin, jasmonic acid, indole-3-acetic acid, etc., were observed during this process (Fig. 1).

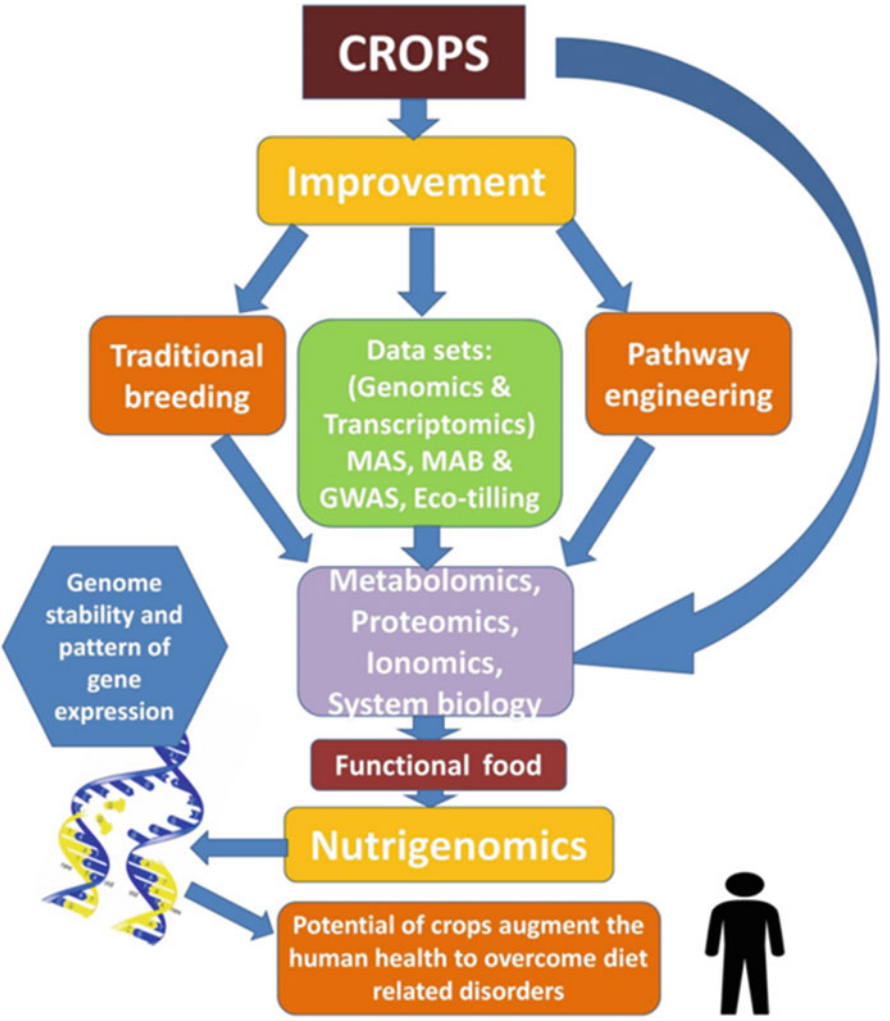


Fig. 1 Nutrigenomics: exploration of crops as functional food for human health and nutrition

6 EcoTILLING

Ecotype targeting induced local lesions in genomes (EcoTILLING) is a variant of TILLING, which examines natural genetic variation in populations and is been successfully utilized in animals and plants to identify the SNPs including rare ones (Barkley and Wnag 2008). A genome-wide association study (GWAS) was employed for elucidating the nutrient concentrations based on seven mineral analyses employing inductively coupled plasma (ICP) spectroscopy (Pratap et al. 2022). A core collection representing accessions collected from 13 different countries identified a total of 6486 high-quality SNPs from the GBS dataset and found 43 marker × trait associations (MTAs) with calcium, iron, potassium, manganese, phosphorous, sulfur, or zinc concentrations by considering 95 cultivated mungbean genotypes collected from the United States Department of Agriculture (USDA) produced in field experiments from 2 consecutive years. The MTAs were distributed in 35 genomic regions, explained on an average of 22% of the variation for each seed nutrient for every year. Most of the genic regions have given important candidate loci for employing in the breeding of new genotypes of mungbean and further in perceptive of genetic control of nutritional traits in the crop. Other identified SNPs in the study will provide an important resource to facilitate MAS for the improvement of nutritional value in mungbean and to assess the cultivars of mungbean (Wu et al. 2020).

7 System Biology

Mungbean is intensively studied; enough data are available on various characters. Data on physical, chemical, food processing, and nutritional properties were obtained from whole mungbean grains and analyzed to reveal the mungbean's potentiality as a food and to understand the research (Dahiya et al., 201). Data revealed that mungbean is a good source of protein (14.6–33.0%) and iron (5.9–7.6 mg/100 g). Grain hardness is linked with fiber content and color is associated with compounds like polyphenols and carotenoids. Physical properties like grain dimensions, sphericity, bulk, porosity, and true density are linked to the moisture percentage of the seed. Antinutrients are phytic acid, tannins, hemagglutinins, and polyphenols. Indicated nutrient contents differ greatly, the reasons for which are not well understood. Grain size and color are associated with different regions and can be used by plant breeders for selection purposes (Saa et al. 2022).

8 Genetic Engineering

Genetic engineering of metabolites is an efficient technology for the development of nutritionally fortified crops (Garg et al. 2018). Protein engineering is considered a potentially important tool in the improvement of proteins in important crops such as mungbean in terms of nutritional and functional attributes (Altenbach et al. 1989; Hoffman et al. 1988; Kim et al. 1990; Utsumi 1992).

8.1 Genetic Engineering of Relevant Biosynthetic Pathways

Mungbean is a well-known source of protein. To increase its bioactivity and nutritional content as a functional food and food additive, lactostatin (IIAEK), a cholesterol-lowering peptide was engineered into mungbean 8Sα globulin, a major storage protein (Gamis et al. 2020). The results indicated that the mutated peptide 8S α globulin has a significant bile acid binding capacity (lowers the cholesterol activity) up to 47.25%. Moreover, superimposed mutant, Mut2, and wild-type, Wt, 3D protein structures showed a 93–97% identity, indicating that the mutant protein's integrity is stable. A similar retention time for wild-type and mutant protein samples was found in ultra-performance liquid chromatography (UPLC)-based assay. Both IIAEK peptide standard and Mut2 digest had comparable baseline peaks, which correspond to the same molecular size based on the data of liquid chromatography-mass spectrometry (LC-MS). A 573.36-Da mass spectrum was seen in Mut2, which indicates that Mut2 8Sα globulin has been successfully mutated and digested to release the bioactive peptide, IIAEK. In vitro bile acid binding capacity showed that the 6-h Wt and 12-h engineered protein (Mut2) digests had the highest cholesterol-lowering activity. Lastly, potential allergenicity was checked in the Allergen Database for Food Safety (ADFS) and the AllerBase database, and the IIAEK peptide matched the Bos d 5 epitopes. This study provides a strong foundation and basis for mungbean nutrition improvement and the development of nutritionally potential cultivars.

The introduction of sulfhydryl groups and disulfide linkage to mungbean 8Sα globulins is attempted by Torio et al. (2011) for the improvement of functional and nutritional qualities. 8Sα globulin or vicilin, the major storage protein of mungbean, was engineered using site-directed mutagenesis to increase the methionine (Met) residues in the protein. Eight Met-rich proteins were engineered and organized to have 2–10 Met residues incorporated in disordered regions, II and IV. The engineered proteins were highly expressed in soluble form in *Escherichia coli*. Their production level of the modified proteins was quantified to be about 30% and was almost the same as that of 8Sα globulin wild type (WT). The modified proteins were in a stable native conformation similar to WT as shown by gel filtration chromatography and also confirmed the greater stability for thermal denaturation, greater emulsifying ability, and emulsion stability, especially the 10-Met protein in comparison to the wild type. Met-rich proteins with 3, 5, and 10 Met residues had 74%, 96%, and 145%, respectively, of the nutritional requirement for Met compared with that of WT, which is 41%. Based on allergenicity prediction programs, WT and all the modified proteins had no allergenic properties.

9 Ionomics

Ionomics is a new platform for understanding the function of the plant ionome during stress and also identifying the genes and regulatory pathways related to mineral accumulation, transportation, and involvement in different molecular mechanisms under normal as well as in stress conditions (Ali et al. 2021). In conjunction with the many expansions in genomics, advances in other "omics" technologies particularly

plant ionomics or ionome profiling have opened up unique opportunities to comprehensively examine the elemental composition and mineral networks of an organism quickly and cost-effectively. These emerging technologies would effectively guide the researchers to enrich the edible parts of grain legumes with bioavailable minerals and enhancers/promoters (Bohra et al. 2015).

Phytic acid (PA) is the storage organic form of phosphorus (P) in the seeds of legumes and cereals. PA is a strong inhibitor against the absorption of nutrients in monogastric animals. The variation of total P (TP) in the seeds of mungbean germplasm and the inheritance property of the seed P compound and phytate contents were studied by Sompong et al. (2010). TP content in seeds of 250 accessions ranged from 2.34 to 5.75 mg/g. The inheritance was studied in the F2 population derived from a cross between two accessions with the lowest (V1658BBR) and highest PA contents (V1141BG). Broad-sense heritability estimates of TP, inorganic P (IP), and phytate P (PhyP) contents were 80.8%, 78.6%, and 80.7%, respectively. The segregation ratio of 9:7 of the F2 population showed that high TP and PhyP were controlled by dominant alleles by two independent loci of major genes indicating duplicated recessive epistasis.

10 Conclusion and Future Perspectives

Pulses are a rich and economical source of proteins and are mainly important in Asian and African vegetarian diets. Among the various pulses, mungbean is an important short-duration pulse crop used in diets across the world, and it has been known to be an excellent source of protein, dietary fiber, minerals, and vitamins. Among the pulses, mungbeans are known for their good digestibility and low flatulence. They are a rich source of phosphorus and pro-vitamin A and are relatively free from anti-nutritional factors. The high protein levels and high lysine/low methionine amino acid profile of mungbean complement the high carbohydrate and low lysine/high methionine content of cereals to form a much balanced amino acid diet. It is the best source of folate as well as iron compared to other pulses. Regular consumption of mungbean seed sprouts in the diet helps in maintaining the beneficial microbial flora in the gut and reduces the absorption of toxic compounds and therefore prevents various metabolic disorders. Recent studies suggest that various bioactive compounds present in mungbean have health-promoting effects on humans. This includes organic acids, flavonoids, and phenolic antioxidants compounds that are known for their effects on detoxification, antihypertension, antipyretic, and anticancer properties. It has great nutritional, agronomic, and pharmacological significance. But the production crop is adversely affected by the cumulative effects of various biotic and abiotic stress factors existing in the environment. Narrow genetic base and poor exploration of available germplasm for valuable traits further restrict the genetic improvement of this crop. The identification and transfer of useful genes from wild species and exotic gene pools into the cultivated environment through various biotechnological and conventional breeding approaches are more promising to develop climate-resilient and nutritionally superior mungbean. It plays important role in agricultural systems and the food sector.

The seeds of mungbean are commonly consumed in processed forms or directly in the form of germinated sprouts in most developing countries. Thus, research on mungbean plays a significant role in nutritional security and soil health on a long-term basis. Identification and development of climate-resilient cultivars should be on priority for the plant breeders involved in the crop improvement of mungbean. The presence of various metabolic compounds in mungbean providing various health benefits has increased the scientific interest in this crop. In the future, mungbean can be explored as a potential source of nutraceuticals and functional foods.

References

Ali S, Tyagi A, Bae H (2021) Ionomic approaches for discovery of novel stress-resilient genes in plants. Int J Mol Sci 22(13):7182

Altenbach SB, Pearson KW, Meeker G, Staraci LC, Sun SS (1989) Enhancement of the methionine content of seed proteins by the expression of a chimeric gene encoding a methionine-rich protein in transgenic plants. Plant Mol Biol 13(5):513–522

Barkley NA, Wang ML (2008) Application of TILLING and EcoTILLING as reverse genetic approaches to elucidate the function of genes in plants and animals. Curr Genomics 9(4):212–226

Bohra A, Sahrawat KL, Kumar S, Joshi R, Parihar AK, Singh U, Singh D, Singh NP (2015) Genetics-and genomics-based interventions for nutritional enhancement of grain legume crops: status and outlook. J Appl Genet 56(2):151–161

Bressani R (1993) Grain quality of common beans. Food Rev Int 9(2):237–297

Bressani R, Elias LG, Garcia-Soto A (1989) Limiting amino acids in raw and processed amaranth grain protein from biological tests. Plant Foods Hum Nutr 39(3):223–234

Cao D, Li H, Yi J, Zhang J, Che H, Cao J, Yang L, Zhu C, Jiang W (2011) Antioxidant properties of the mungbean flavonoids on alleviating heat stress. PLoS One 6(6):e21071

Cozzolino SMF, Cominetti C (2013) Biochemical and physiological bases of nutrition in different stages of life in health and disease. Monole, Sao Paulo

Cruz JC, Tseng HC, Goldman JA, Shih H, Tsai LH (2003) Aberrant Cdk5 activation by p25 triggers pathological events leading to neurodegeneration and neurofibrillary tangles. Neuron 40(3):471–483

Dahiya PK, Linnemann AR, Van Boekel MAJS, Khetarpaul N, Grewal RB, Nout MJR (2015a) Mungbean: technological and nutritional potential. Crit Rev Food Sci Nutr 55(5):670–688

Dahiya S, Sarkar O, Swamy YV, Mohan SV (2015b) Acidogenic fermentation of food waste for volatile fatty acid production with co-generation of biohydrogen. Bioresour Technol 182:103–113

Datta S, Mahfooz S, Singh P, Choudhary AK, Singh F, Kumar S (2010) Cross-genera amplification of informative microsatellite markers from common bean and lentil for the assessment of genetic diversity in pigeonpea. Physiol Mol Biol Plants 16(2):123–134

Dauncey MJ (2014) Overview of nutrigenomics and epigenomics: mechanisms and relevance. In Proceedings of FACTA avian nutrigenomics course, Campinas SP, Brazil, 27–28 May 2014

Di Renzo L, Gualtieri P, Romano L, Marrone G, Noce A, Pujia A, Perrone MA, Aiello V, Colica C, De Lorenzo A (2019) Role of personalized nutrition in chronic-degenerative diseases. Nutrients 11(8):1707

Egawa Y, Tomooka N (1994) Phylogenetic differentiation of Vigna species in Asia. JIRCAS Int Symp Ser 2:112–120

Fialho AM, Moreira LM, Granja AT, Popescu AO, Hoffmann K, Sá-Correia I (2008) Occurrence, production, and applications of gellan: current state and perspectives. Appl Microbiol Biotechnol 79(6):889–900

Gamis MC, Uy LY, Laurena A, Hurtada W, Torio MA (2020) Protein engineering of mungbean (*Vigna radiata* (L.) Wilczek) 8Sα globulin with lactostatin. Appl Sci 10(24):8787

Gan RY, Lui WY, Wu K, Chan CL, Dai SH, Sui ZQ, Corke H (2017) Bioactive compounds and bioactivities of germinated edible seeds and sprouts: an updated review. Trends Food Sci Technol 59:1–14

Garg, R. and Jain, M., 2013. Transcriptome analyses in legumes: a resource for functional genomics Plant Genome, 6(3), pp plantgenome2013-04

Garg M, Sharma N, Sharma S, Kapoor P, Kumar A, Chunduri V, Arora P (2018) Biofortified crops generated by breeding, agronomy, and transgenic approaches are improving lives of millions of people around the world. Front Nutr 5:12

Gupta PK, Langridge P, Mir RR (2010) Marker-assisted wheat breeding: present status and future possibilities. Mol Breed 26(2):145–161

Habibullah AM, Abbas M, Shah HU (2007) Proximate and mineral composition of mungbean. Sarhad J Agric 23:462–466

Hamwieh A, Udupa SM, Sarker A, Jung C, Baum M (2009) Development of new microsatellite markers and their application in the analysis of genetic diversity in lentils. Breed Sci 59(1):77–86

Hanumantha Rao B, Nair RM, Nayyar H (2016) Salinity and high temperature tolerance in mungbean [*Vigna radiata* (L.) Wilczek] from a physiological perspective. *Front*. Plant Sci 7:957

Hashiguchi A, Zhu W, Tian J, Komatsu S (2017) Datasets of mungbean proteins and metabolites from four different cultivars. Data Brief 13:703–706

Hoffman LM, Donaldson DD, Herman EM (1988) A modified storage protein is synthesized, processed, and degraded in the seeds of transgenic plants. Plant Mol Biol 11(6):717–729

Jain KK, Kumar A, Shankar A, Pandey D, Chaudhary B, Sharma KK (2020) De novo transcriptome assembly and protein profiling of copper-induced lignocellulolytic fungus Ganodermalucidum MDU-7 reveals genes involved in lignocellulose degradation and terpenoid biosynthetic pathways. Genomics 112(1):184–198

Kang YJ, Kim SK, Kim MY, Lestari P, Kim KH, Ha BK, Jun TH, Hwang WJ, Lee T, Lee J, Shim S (2014) Genome sequence of mungbean and insights into evolution within *Vigna* species. Nat Commun 5(1):1–9

Kim CS, Kamiya S, Kanamori J, Utsumi S, KITO, M. (1990) High-level expression, purification and functional properties of soybean proglycinin from *Escherichia coli*. Agric Biol Chem 54(6):1543–1550

Kole C, Abbott AG (2011) International Workshop on Phytomedimics and Nutriomics and ICPN organized by the PD at San Diego, CA on 17 January, 2011 (www.intlpag.org)

Kumar J, Choudhary AK, Solanki RK, Pratap A (2011) Towards marker-assisted selection in pulses: a review. Plant Breed 130(3):297–313

Lakhanpaul S, Babu CR (1991) Genome size and evolution in Asiatic *Vigna Savi*. In: Symposium on Grain Legumes, New Delhi, pp. 47–57

Lee JH, Jeon JK, Kim SG, Kim SH, Chun T, Imm JY (2011) Comparative analyses of total phenols, flavonoids, saponins and antioxidant activity in yellow soy beans and mungbeans. Int J Food Sci Technol 46(12):2513–2519

Liener IE (1994) Implications of antinutritional components in soybean foods. Crit Rev Food Sci Nutr 34(1):31–67

Liener IE (1995) Possible adverse effects of soybean anticarcinogens. J Nutr 125(suppl_3):744S–750S

Liener IE, KakADE ML (1980) Protease inhibitors. Toxic Constituents Plant Foodstuffs 6:7

Liu B, Qian SB (2011) Translational regulation in nutrigenomics. Am Soc Nutr 2:511–519

Lu Y, Chang X, Guo X (2019) Dynamic changes of ascorbic acid, phenolics biosynthesis and antioxidant activities in mungbeans (*Vigna radiata*) until maturation. Plan Theory 8(3):75

Moghadam MB, Vazan S, Darvishi B, Golzardi F, Farahani ME (2011) Effect of mungbean (Vigna radiate) living mulch on density and dry weight of weeds in corn (Zea mays) field. Commun. Agric. Appl. Biol. Sci. 76, 555–559

Mubarak AE (2005) Nutritional composition and antinutritional factors of mungbean seeds (*Phaseolus aureus*) as affected by some home traditional processes. Food Chem 89(4):489–495

Nair RM, Pandey AK, War AR, Hanumantharao B, Shwe T, Alam AKMM, Pratap A, Malik SR, Karimi R, Mbeyagala EK, Douglas CA (2019) Biotic and abiotic constraints in mungbean production—progress in genetic improvement. Front Plant Sci 10:1340

Pandey AK, Burlakoti RR, Kenyon L, Nair RM (2018) Perspectives and challenges for sustainable management of fungal diseases of mungbean [*Vigna radiata* (L.) R. Wilczek var. radiata]: a review. Front Environ Sci 6:53

Parida A, Raina S, Narayan R (1990) Quantitative DNA variation between and within chromosome complements of *Vigna* species (Fabaceae). Genetica 82:125–133

Peng X, Zheng Z, Cheng KW, Shan F, Ren GX, Chen F, Wang M (2008) Inhibitory effect of mungbean extract and its constituentsvitexin and isovitexin on the formation of advanced glycationendproducts. Food Chem 106(2):475–481

Pratap A, Dhaliwal I, Singh CM, Mahalingam A, Manivannan N, Basavaraja T, Rathore M, Singh Y, Tiwari P, Yadav S, Tripathi A (2022) Biofortification of Mungbean. In: Kumar S, Dikshit HK, Mishra GP, Singh A (eds) Biofortification of staple crops. Springer, Singapore, pp 295–333

Reddy MRK, Rathour R, Kumar N, Katoch P, Sharma TR (2010) Cross-genera legume SSR markers for analysis of genetic diversity in Lens species. Plant Breed 129(5):514–518

Ronteltap A, Van Trijp JCM, Renes RJ (2008) Consumer acceptance of nutrigenomics-based personalised nutrition. Br J Nutr 101(1):132–144

Saa RW, Fombang Nig E, Radha C, Ndjantou EB, Njintang Yanou N (2022) Effect of soaking, germination, and roasting on the proximate composition, antinutrient content, and some physico-chemical properties of defatted Moringa oleifera seed flour. J Food Process Preserv 46(3):e16329

Sakai H, Naito K, Takahashi Y, Sato T, Yamamoto T, Muto I, Itoh T, Tomooka N (2016) The Vigna Genome Server, 'Vig GS': a genomic knowledge base of the genus *Vigna* based on high-quality, annotated genome sequence of the Azuki Bean, *Vigna angularis* (Willd.) Ohwi & Ohashi. Plant Cell Physiol 57(1):e2–e2

Samtiya M, Aluko RE, Dhewa T (2020) Plant food anti-nutritional factors and their reduction strategies: an overview. Food Prod Process Nutr 2(1):1–14

Saxena RK, Prathima C, Saxena KB, Hoisington DA, Singh NK, Varshney RK (2010) Novel SSR markers for polymorphism detection in pigeonpea (*Cajanus* spp.). Plant Breed 129(2):142–148

Shi Z, Yao Y, Zhu Y, Ren G (2016) Nutritional composition and antioxidant activity of twenty mungbean cultivars in China. Crop J 4(5):398–406

Shrestha S, van't Hag L, Haritos VS, Dhital S (2022) Lentil and Mungbean protein isolates: processing, functional properties, and potential food applications. Food Hydrocoll:108142

Singh R (2013) Development of iron and zinc enriched mungbean (Vigna radiata L.) cultivars with agronomic traits in consideration. Wageningen University and Research

Singh B, Singh JP, Kaur A, Singh N (2017a) Phenolic composition and antioxidant potential of grain legume seeds: a review. Food Res Int 101:1–16

Singh B, Singh JP, Singh N, Kaur A (2017b) Saponins in pulses and their health promoting activities: a review. Food Chem 233:540–549

Sompong U, Kaewprasit C, Nakasathien S, Srinives P (2010) Inheritance of seed phytate in mungbean (*Vigna radiata* (L.) Wilczek). Euphytica 171(3):389–396

Torio MAO, Adachi M, Garcia RN, Prak K, Maruyama N, Utsumi S, Tecson-Mendoza EM (2011) Effects of engineered methionine in the 8Sα globulin of mungbean on its physicochemical and functional properties and potential nutritional quality. Food Res Int 44(9):2984–2990

Utsumi S (1992) Plant food protein engineering. Adv Food Nutr Res 36:89–208

Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27(9):522–530

Wang H, Guo X, Li Q, Lu Y, Huang W, Zhang F, Chen L, Liu RH, Yan S (2020) Integrated transcriptomic and metabolic framework for carbon metabolism and plant hormones regulation in Vignaradiata during post-germination seedling growth. Sci Rep 10(1):1–11

Wu X, Islam AF, Limpot N, Mackasmiel L, Mierzwa J, Cortés AJ, Blair MW (2020) Genome-wide Snp identification and association mapping for seed mineral concentration in mungbean (*Vigna radiata* L.). Front Genet 11:656

Yi-Shen Z, Shuai S, FitzGerald R (2018) Mungbean proteins and peptides: nutritional, functional and bioactive properties. Food Nutr Res 62

Rice Bean: A Neglected and Underutilized Food Crop Emerges as a Repertory of Micronutrients Essential for Sustainable Food and Nutritional Security

Tanushri Kaul, Jyotsna Bharti, Rachana Verma, Puja Chakraborty, Arulprakash Thangaraj, Mamta Nehra, Sonia Khan Sony, Khaled Fathy, Rashmi Kaul, and Murugesh Easwaran

Contents

Abstract

Rice bean (*Vigna umbellata* syn. *Phaseolus calcarata* (Roxb.); *Azuki umbellata* (Thunb.) Owhi and Ohashi) is an obscure crop that is underexposed to the scientific investigation as yet. However, the crop has the essence of a potential crop, and it is considered a food and nutrition security grain for farmers of Southeast Asia, where rice bean is grown by small farm holders on marginal land with minimum agricultural inputs. Due to its potent nutrient profile, it is grown and diversified since its origin from Indo-China region; however, the crop till now has not received the requisite attention. The agronomic disadvantage like low palatability and indeterminate growth pattern with delayed flowering hampers the productivity and quality of the crop product and limits its spread, utilization, and commercialization. However, there has been a renewed interest in this underutilized crop, because it is a great genetic source of biotic and abiotic stress tolerance. Recent omics investigations and the literature available have highlighted the genes involved in the agronomically negative transactional traits.

T. Kaul (✉) · J. Bharti · R. Verma · P. Chakraborty · A. Thangaraj · M. Nehra · S. K. Sony · K. Fathy · R. Kaul · M. Easwaran
Nutritional Improvement of Crops (NIC) Group, International Centre for Genetic Engineering and Biotechnology (ICGGEB), New Delhi, India
e-mail: tanushri@icgeb.res.in; kaultanushri3@gmail.com

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_32

The identification of vital target genes is a prerequisite for trait improvement. Crop improvement and the identification of new target crops at the verge of population explosion, climatic, and environmental changes are crucial. The present endeavor has tried to highlight the importance of rice bean and recognize it as a crop for the future since staple crops are facing challenges for further improvement. We have discussed the work done in this direction along with the recent revelations and the need for genetic improvement for developing the rice bean as robust, high quality, and yielding crop.

Keywords

Rice bean · Underutilized crop · Nutrition · Genomic resource · Omics · Crop improvement

1 Introduction

A versatile, underutilized, less well-known legume known as the rice bean (*Vigna umbellata* syn. *Phaseolus calcarata* (Roxb.); *Azuki umbellata* (Thunb.) Owhi and Ohashi) has the potential to provide the rising human population with nutritious feed in the near future (Bhat and Karim 2009). It is a member of the Fabaceae (Leguminosae) family (Bhardwaj et al. 2021; Dahipahle et al. 2017) (Fig. 1). Therefore, it has the innate ability to assimilate biological nitrogen to restore soil fertility and serve as a substitute for intercropping. It is regarded as a resource for nutritious flour, dietary needs, and livestock feeding and fodder (Dahipahle et al. 2017; Khanal et al. 2009). (Tripathi et al. 2021). Although it holds such worthwhile agricultural capabilities and nutritional brilliancy, rice bean still has been abandoned by breeders. Attributions such as delay in flowering (Kaul et al. 2019a; Joshi et al. 2007; Takahashi et al. 2015) and unpalatableness (Basu and Scholten 2012) are two major constraints associated with cultivated varieties. More is concerned with the inadequate information on its nutritious advantage and the presence of anti-nutrients (Katoch 2020). Minimal advancement has been accomplished in achieving potentiality of rice bean due to the main matters associated to late induction of flowering (Joshi et al. 2007), unpalatableness (Kaul et al. 2019a), hard and coarse type of grain (Andersen 2012), shattering sensitivity (Parker et al. 2021), disease resilience (Pandiyan et al. 2008), and the presence of anti-nutrients (Bajaj 2014). Several kinds of literature evince rice beans' resiliency towards adverse environments consisting of some familiar biotic and abiotic stresses along with abilities to tolerate metal toxicity in the soil. *Vigna umbellata* has been identified by the Food Security, Rice Bean Research in India and Nepal (FOSRIN) network as one of the impending crops for the nutrified future (Andersen 2007; Basavaprabhu et al. 2013). Originated from Southern and Southwestern Asia (Bisht and Singh 2013), it had been adapted in Indo-China area (Doi et al. 2002) and became a conventional crop in East, Southeast, and South Asia (Tomooka et al. 2002; Seehalak et al. 2006). It is mostly cultivated in the belt of Indonesia, Vietnam, Myanmar, Bhutan, Southern China, Laos, Northern Thailand, East Timor, and India

Fig. 1 Life cycle of rice bean: (**a**) seedling stage; (**b**) reproductive stage: vine legume plants with yellow flowers and small edible beans; (**c**) inflorescence having flowers in different stages; and (**d**) variability of seed coat color and seed size (diameter of each unit is 4 cm) in rice bean. The vegetative parts can be fed to livestock as fresh or made into hay. Grain is generally used as dal/boiled soup for human consumption. Grain foliage is used as livestock fodder and manure

(Tian et al. 2013; Iangrai et al. 2017; Pattanayak et al. 2018). The wild type of rice bean, *V. umbellata* varieties *gracilis*, inhabiting the zone from Himalayan and Central Land of China to Malaysia (Seehalak et al. 2006; Rejaul et al. 2016), is presumed as the pioneer of today's cultivated crop. Although wild types exhibit intermediacy in qualities such as small seeds, free branching, and sensitivity towards photoperiod as compared to the cultivated varieties, till now, only the northern east area of India grow such wild-type intermediate varieties (Iangrai et al. 2017). Research on related members like *V. mungo* revealed the coevolution of this species during an individual domestication event in Thailand or Myanmar. Conspecific to the cultivated *V. umbellata* grown in South and Southeast Asia, this species is thought to have arisen from *V. umbellata* var. gracilis (Tomooka et al. 1991, 2002; Bisht et al. 2005; Seehalak et al. 2006). Rice bean has been embraced as a cover crop in places like Ghana, Sri Lanka, Indonesia, Jamaica, Haiti, Fiji, and Mexico, but it is only in limited cultivation

in West Indies, Australia, Africa, Brazil, the United States, and Honduras (Wang et al. 2015; Khadka and Acharya 2009; Rajerison 2006; Burkill 1953) (Fig. 2). The location of the crop is mostly determined by the necessary growth factors; the rice bean thrives in regions with an annual rainfall of between 1000 and 1500 mm, although it also tolerates dry seasons to a reasonable extent. This crop can grow in the temperatures ranging between 18 °C and 30 °C, but it also can withstand cold temperatures as low as 10 °C with a vulnerability to more lower temperatures like frost, as well as temperatures as warmer as 40 °C. (Pattanayak et al. 2019). The screening of rice bean's germplasm for unexplored genomes may surface new approaches for unearthing neoteric traits and linked genes (Sarma et al. 1995). These agricultural endeavors favor generating robust and climate-resilient crops against the backdrop of adverse environments. The current circumstances of world hunger and the tendency of developing lifestyle-related severe disorders demand improvement and domestication of unexplored crops like rice bean to create unique sustainable landraces with increased productivity and lower levels of anti-nutrients. Rice bean may be enlisted as one of the crucial futuristic crops. In order to accelerate its domestication process, the crop necessitates more focused research. In comparison to other economically significant legumes including soybean, common bean, azuki bean, and mung bean, the absence of polymorphic molecular markers in rice bean has lagged in genomic investigations. So far, only a few papers have indicated intra- and inter-species molecular diversification of *V. umbellata* via application of simple sequence repeat (SSR), amplified fragment length polymorphism (AFLP), and random

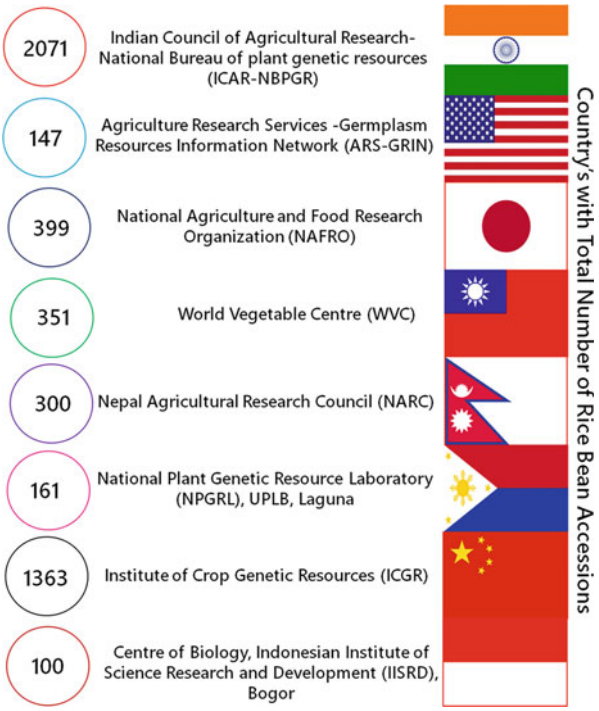


Fig. 2 Diagrammatic representation depicting the worldwide distribution of rice bean accessions in diverse agroecological microhabitats

amplified polymorphic DNA (RAPD) markers (Iangrai et al. 2017; Thakur et al. 2017; Tian et al. 2013; Bajracharya et al. 2008; Muthusamy et al. 2008). An in-depth draft genome assembly of rice bean will provide information on nucleotide polymorphism existing in the population, which will be turned into the resource of polymorphic SSRs applicable in determining how the marker linked to particular trait. The molecular breeding strategy for crop advancement can be efficiently regulated by the identification of genome-wide relevant trait-related markers, particularly markers linked to quantitative traits (Muthusamy et al. 2008). Landraces from various geographical areas must be included in order to fully represent the genetic variety. SSR markers addressed the greatest variation in an examination of 112 regionally collected different rice bean varieties from India and Nepal using 35 azuki bean's polymorphism markers (Bajracharya et al. 2008). To analyze the diversity indices for the target crop's genotype, polymorphic markers created for the orthologous crop can be used for the less-known sister crops. It is important to note that the first comprehensive SSR analysis, which included 84 wild and 388 cultivated accessions from 16 different geographic zones in Asia, revealed that the genetic diversity is highest in accessions from Nepal, Myanmar, Vietnam, and India (Tian et al. 2013). Among 65 accessions investigated using a cluster of 28 SSR markers, accessions were collected from Japan, Thailand, Korea, and China with significant diversified genetic characteristics. It is crucial to carefully examine the rice bean germplasm in order to benefit from the diversity that is intrinsically present there and to minimize the risks associated with farmers giving up the crop. Therefore, more emphasis on genetically governed key characteristics of rice beans is highly required for its adaptation, like taste, latency, and early floral induction to make this crop more feasible. This kind of approach permits the study of evolutionary components and the effects of domestication and breeding on rice bean molecular function (Fig. 1).

What necessitate the exploration of the underutilized crops? Will the endeavors be advantageous to resolve the problem of yield loss, biotic/abiotic stress vulnerability, and malnourishment? Invariably, these grave questions are relevant during a time when staple food crops are facing numerous challenges (Mayes et al. 2012; Andersen et al. 2009); moreover, the situation will escalate with the discrepancy in the current high population growth rate and low food productivity. Thus, with the upsurge to provide safe and healthy nutrition for underprivileged and deprived populations, food grain productivity must be increased urgently in order to feed the rapidly growing population. Considerably, the underutilized crops are more resilient to the changing environment or biotic/abiotic stress and well equipped to dwell on marginal lands with minimal investments of capital and agricultural products. Thus, the underutilized, neglected, or non-staple food crops are generally accepted alternatives to meet the Sustainable Development Goal (SDG).

2 Nutritional Composition of Rice Bean

Compared to other commercially grown crops, rice beans have a considerable amount of nutritional content and various health advantages. Furthermore, it holds tremendous potential to mitigate global micronutrient deficiency among the population. In South

Asian nations, rice bean is often eaten as a boiled soup or dal, as well as whole grain cooked by using dry heat and hot air, combined with wheat flour or chickpea. They provide a substantial quantity of nutritional fiber, tannin, slowly absorbed carbohydrates, significant minerals, and important amino acids. Rice bean has more iron (Fe), calcium (Ca), zinc (Zn), potassium (K), thiamine, riboflavin, and niacin compared to other leguminous crops. Moreover, it has a lower glycemic index (Katoch et al. 2014; Saharan et al. 2002). A large variety of bioactive compounds contain those that protect the liver, fight inflammation, lower blood pressure, improve immunity, fight cancer, fight infections and fungal growth, fight diabetes, inhibit HIV-1, and fight health hazards (Wei et al. 2015). In particular, ischemic heart disease, diabetes mellitus-type 2, colorectal cancer, and other metabolic ailments can be curbed and/or prevented by the bioactive phytochemical profile of rice beans (Tharanathan and Mahadevamma 2003). Rice bean is becoming an excellent choice for underdeveloped and developing countries because the nutritive profile holds huge health and nutritive benefits (Dhillon and Tanwar 2018; Katoch 2013; Parvathi and Kumar 2006). Digestible and non-digestible carbohydrates are the primary component of any legume (Priyadarshini et al. 2021). Rice bean contains more than 58% carbohydrate and up to 72% carbohydrate (Katoch 2013; Buergelt et al. 2009; Sadana et al. 2006), from 3.60% to ~5.56% of crude fiber (Bajaj 2014; Buergelt et al. 2009), while ~9 g per 100 g of acid detergent fiber (ADF) and ~13.0 g per 100 g of neutral detergent fiber (NDF) are present (Katoch 2013). Rice bean has 5–6 g/100 g of soluble sugars, 4.7–5.3 g/100 g of nonreducing sugars, and comparatively low starch (50–55 g/100 g) compared to other beans (Saharan et al. 2002). Rice bean has fewer oligosaccharides like raffinose, which ranges from 1.56% to 2.58%, while verbascose and stachyose are present in a percentage of 0.85–1.23 and 0.94–1.88, respectively. Food rich in raffinose and stachyose such as soybean, sword bean, and lima bean causes flatulence in the human body (Katoch 2013).

Rice bean has a total protein level ranging from 14% to 26% with better protein digestibility (86.1–89%) in vitro as compared to other legume crops (Bajaj 2014; Buergelt et al. 2009). Katoch (2013) reported rice beans have a greater amino acid profile, specifically tryptophan and methionine, whereas black and green gram have equivalent tyrosine, lysine, and valine levels. The reported fat content of 1.92–3.42% is also significantly lower than that of legumes (Bepary et al. 2017; Buergelt et al. 2009; Sadana et al. 2006). Rice bean can provide a resource for unsaturated fatty acids such as oleic acid and linoleic acid with a percentage of 15.6–17.91 and 17.24–18.98, respectively, while linolenic acid, stearic acid, and palmitic acid with a percentage of 39.89–44.36, 4.36–5.87, and 14.23–16.88, respectively (Katoch 2013; Pugalenthi et al. 2004). In rice bean, Katoch (2013) discovered 3.48–4.26 mg/100 g niacin and 15.33–29.00 mg/100 g ascorbic acid. Interestingly, phytic acid, a key anti-nutrient feature in legume grains, is present in trace amounts of 0.20–2.27% and is comparatively less than in other crops (Bepary et al. 2017; Bajaj et al. 2014). However, the anti-nutrient qualities, such as activities of phytic acid, polyphenols, trypsin inhibitors, and saponins, can be reduced by soaking and pressure cooking. As a result, its nutritional content may offer a hopeful addition of healthy and inexpensive value to impoverished and advanced nations in order to counter the food crisis (Dhillon and Tanwar 2018).

3 Problems with the Commercial Use of Rice Beans

Despite having health benefits, rice beans are considered an underdeveloped crop; however, it is difficult to pinpoint what characteristics make a crop underutilized: they are frequently linked to the cultural heritage of their region of origin; there is inadequate evidence of their farming and use; they are customized to specific marginal land and agroecological microhabitats; and there is no formal document for seed distribution mechanisms. Its anti-nutritional components, including phenolic compounds, alkaloids, phytate, enzyme inhibitors, late flowering, prolonged cooking time, inferior processing, and tough seed coatings, all contribute to its poorer usage. Additionally, its anti-nutritional contents, notably phenolic compounds, tannins, phytate, enzyme inhibitors, late flowering, extended cooking time, poor processing methods, and hard seed coats, contributes to its less utilization. Thereby, farmers are less eager to spend their time and resources on the expansion and production of underutilized crops (Nnamani et al. 2017; Popoola et al. 2020). In addition, rice bean is mostly grown by rural farmers who lack the resources, i.e., high-input farming practices required to raise essential crops (Conti et al. 2019). Additionally, their production has been limited by low market prices, low demand, a lack of purchasers, an inadequacy of improved varieties, and insufficient funding (Khan et al. 2021). In addition, light sensitivity and indeterminate vegetative developments due to delayed flowering have lost market attraction compared to crops like chickpeas, pigeon peas, lentils, peas, and black and green grams. Rice bean verities with prolonged vegetative development periods seem unpleasant, and the subsequent cropping is hampered by this prolonged period. Also, the tough and coarse grains limit daily consumption even after boiling (Andersen 2012). Early flowering may be accelerated by manipulating genes involved with the flowering pathway (González et al. 2016; Dhanasekar and Reddy 2014; Joshi et al. 2007). Furthermore, it is reported that rice beans may grow in a variety of climates. However, their cultivation is restricted by their need for moderate temperatures and rainfall and their sensitivity to prolonged exposure to harsh settings (Noda 1951). Rice bean cultivation is reduced in the cropping system due to the crop's rigid reaction to the mechanized cultivation system and lower harvest index, which contrast with other economically relevant pulses like lentils, peas, and soybean (Jayasundara 2015). Due to these serious issues, there are no legal channels for trade; still, it is locally marketed in Thailand, Nepal, India, and Myanmar through an unorganized trade.

4 Omics Approach: Identifying Novel Genes Associated with Stress Resistance and Nutritional Improvement

Due to the limited geographical distribution in congruence with restricted marketing channels and inadequate commercialization, rice bean has been unexplored for their underlying genetic richness. Thus, unraveling the rice bean' genome is critical, but it could provide ideal candidate genes for stress resilience and an extraordinary range of nutrients within the genus *Vigna* (*V. umbellata*) (Rana et al. 2014). Thereby, a major

contribution has been made by Kaul et al. (2019a) with a 414 MBP genome draft assembly, i.e., NCBI.SRA.SRP132447 consisting of high confidence, identified 31,276 genes via analysis of 15,521 scaffolds. Primarily, it is critical to cover the whole genome. Functional coverage of almost 96.08% was attained via coverage of around 30X reads developed through Illumina along with the PacBio approach (Kaul et al. 2019a, 2022). The assembled genome indicated the closest relation of *Vigna umbellata* genome with *V. angularis*, followed by another closely related associate present in same genus, i.e., *Vigna radiata* and *Vigna unguiculata*. Collinearity block mapping, an alignment technique to compare crop genome assembly to other related plants' genomes, is employed and aligns the rice bean draft assembly with 13 complete genomes and 18 partial genomes available for legume crops. It revealed peculiar information about rice beans. The whole coding sequence (CDS) alignment also provided critically relevant findings. Moreover, with the application of LCB, i.e., locally collinear alignment block clusters, aligning with 17 medicinally relevant plant genomes incorporated in the National Institute of General Medical Sciences-NIH database, are able to decipher the 18,000 potential medicinally pertinent genes. In conclusion, the endeavors encompassing comparative genomics studies have highlighted the fundamental symplesiomorphic traits that have assisted in establishing the origin and relation of rice beans related to their genetic and functional lineage. Considerably, the major contribution of this study is the identification of neoteric palatability and late-flowering related genes and deciphering the position of a few of them in the mitochondria and chloroplast genomes. These genes are involved in functionally diverse metabolic pathways for the regulation of flavor, abiotic/biotic stress, disease resistance, photoperiod responsiveness, and more. The material and data have been made available at http://www.nicg.in to advance molecular breeding studies to produce rice beans as a potential resource. Kaul et al. (2022) reported the existence of genes that responded to stress from a variety of families, including stress-enhanced protein 1 (SEP-1), universal stress protein PHOS32 precursor (PHOS32), heat shock transcription factor HSF-02 (GMHSF-02), stress-responsive alpha-beta barrel domain-containing protein (GSU2970), and stress-enhanced protein 2 (SEP-2) in rice bean (Kaul et al. 2022). Targeting photoperiod-independent early flowering genes like the early flowering 3 (Elf3) gene, FLC-transcription factors, flowering locus T1 (Rft1), determinant stem 1 (DT1), ethylene-responsive transcription factor tiny (TINY), and Dead/dead box helicase domain-containing protein (PIE1) can increase yield by lowering vegetative and indeterminate growth. In India, the Nutritional Improvement of Crops (NIC) Group at ICGEB, New Delhi, has pioneered molecular studies in rice beans to add features including determinate habit, early flowering, and palatability using a clustered regularly interspaced short palindromic (CRIPSR)-based genome editing system (Fig. 3).

The advances in sequencing technology have made genomic exploration less challenging, which contributes primarily toward the visible paradigm shift in agriculture. Conventionally, till now, quantitative trait locus (QTL) mapping has harnessed the uncloaked capabilities of these underutilized crops and characters related to domestication via inter- and intraspecific mapping (Isemura et al. 2010). Rice bean projects resistance capability against mung bean yellow mosaic virus

Fig. 3 Schematic representation depicting the genome editing approach to target the palatability and late flowering genes to resolve the issues with the rice bean. Recent innovations like CRISPR/Cas9 could accelerate the breeding process

(Pandiyan et al. 2008, 2010; Sudha et al. 2013), pests actively affect storage conditions such as bruchid beetles (*Callosobruchus* spp.) (Tomooka et al. 2000; Kashiwaba et al. 2003; Somta et al. 2006) and bacterial disease like leaf spot (Arora et al. 1980). Generally, the underused crops are efficiently habituated to a variety of soil types. Specifically, the inherent capability of rice beans to secrete organic acids like citric acid and others curb the acidity of the soil and aluminum accumulation (Fan et al. 2014). MATE has identified resistance power to aluminum toxicity in rice beans, i.e., multidrug and toxic compound extrusion family organic acid efflux transporters such as VuMATE1 and VuMATE2 (Yang et al. 2006; Liu et al. 2018). Furthermore, the absorption of micronutrients like iron and zinc is reduced due to the prevalence of anti-nutrients such as polyphenols, saponins, tannin, phytic acid (PA), hemagglutination, and trypsin inhibitors. Thereby, the identified followed by functionally characterized micronutrient transporters can be targeted for the improvement of nutrition in rice beans such as iron-phytosiderophore transporter yellow stripe 1 (YS1), iron-regulated transporter-1(IRT1), constitutive photomorphogenic 1 (COP1), ferric reduction oxidase 2 (FRO2), nicotinamide synthase 1 (NAS1), and natural resistance-associated macrophage protein (NRAMP). In general, the PA residues form a complex with the micronutrient and thus make it inaccessible for absorption in the animal body too. Therefore, targeting genes such as phospholipase D (PLD) and inositol triphosphate kinase (ITPK6) can lead to a lower phytic acid concentration in rice bean grain. Likewise, the main reason for the low palatability index of rice beans is the presence of taste deterrents or pungent taste developers, for instance, polyphenols and tannins. Thereby, genes like N-(5-phosphoribosyl) anthranilate isomerase – a chloroplastic isoform, liquiritigenin 2-*O*-methyltransferase-like (ILMT), chalcone synthase 17-like (CHS-17), leucoanthocyanidin dioxygenase (LDOX), and spermidine hydroxycinnamoyl transferase-like (SHT) can be targeted due to their involvement in biosynthesis pathways of polyphenol and tannin. In accordance with this, the application of genome editing may lead the way to

understanding the functional characterization of genes connected to critical agricultural factors determining crop habituation (Schenke and Cai 2020; Zafar et al. 2020).

5 Crop Improvement

If the underutilized crops are as superior as they are represented and promoted as having agro-economic, stress tolerance, and nutritious advantages, why then don't they reproduce on their own? Will they suffer if agriculture is increasingly commercialized and the food trade is more international? These barriers can be overcome using the traditional and molecular methods outlined below. These crops are frequently discussed in cross-disciplinary stakeholder/partnership efforts in agricultural progress (Padulosi and Hoeschle-Zeledon 2004, 2008), and as a result, they are covered under post-green revolution development policies. Both producers and consumers are unenlightened of rice beans' tremendous potential to enhance food and nutritional advantages, so, emphasizing its marketing and utility are highly significant. For novel varieties and flexibility in seed selection, old variety material should be exploited. For crop improvement and propagation, it demands a focus on breeding and on-farm trials. Seed providers should promote seed distribution and accessibility easily. By educating larger audiences about rice bean seed production in native places and handing out grains at trade fairs, access can be increased. In addition, quality assessment is vital (regarding germination and resilience to disease). Furthermore, timely circularization of information regarding the adaptability and technologies related to better production of rice beans is necessary. Alliance with the nongovernmental organization could be the marketing strategy to popularize the rice bean at chains of food hubs (Sthapit et al. 2010). Nutritional, chemical, and sensory analyses must be carried out at several intervals to gain data from various marketing surveys. These methods may create a unique hub for global consumption, achieving security in nutritious food and reducing malnutrition. Moreover, recent advances and implementation of high-throughput next-generation DNA sequencing approaches, like Illumina and Pac bio, have massively furnished the underlying potential of rice bean genomes with high confidence and accurately aligned data. Kang et al. (2014) released the draft genome of *V. radiata* (~genome size of almost 459 MB) with three genome assemblies and one sequencing read. They showed the draft genome of *V. angularis* with a genome size of ~455 MB with five assembled genomes and two sequencing reads by Kang et al. (2015). Similar findings involving *V. unguiculata* with an approximate genome size of 607 MB were reported by Muoz-Amatria et al. in 2017. The researchers developed the draft genome using two assembled genomes. Using next-generation sequencing (NGS) on the rice bean genome, Chen et al. (2016) were able to locate 3011 possible genetic/molecular SSR markers. But before drafting a genome to build confidence and partially annotate it, it is essential to perform an analysis of comparative genomes with the target genome. Kaul et al. (2019a, 2022) mapped the whole genome of rice beans using the anchoring elements of the functional and genetic orthologous sequences of *V. radiata*, *V. angularis*, and *V. unguiculata* as a basis for comparable functional properties. The blueprint of the draft genome, identification of genes, and translational annotation for the target *Vigna*

umbellata can all be designed using the orthologous sequences as a reference. The probable protein-coding genes can also be found from the assembled sequence data utilizing the process, namely MAKER v2.31.9. (Campbell et al. 2014). It results in annotating genes via quality-dependent evidence. These eventually lead to the identification of the target gene or genes for trait modification and raise the market value of the neglected target crop (Kaul et al. 2019a). Genetic engineering techniques must be used to improve traits (Tabassum et al. 2021). Among the several genome editing techniques, CRISPR-dependent toolkits are the most effective, highly specific, and simple to generate editing techniques, which have recently gained attention in the researchers' community. The aforementioned method updates the targeted genome modification appropriately and suitably (Zhang et al. 2018; Ku and Ha 2020; Kaul et al. 2019b, 2020). These have been used extensively to discover traits and develop high-yield crops (Wang et al. 2018, 2019). Similar to this, rice bean characteristics such as less palatability, late regulation of flowering, and presence of anti-nutrients, alongside photoperiod sensitiveness, can be altered using CRISPR/Cas-dependent methods. As a result, it might be a ground plan to fix and eliminate the problems linked to the domestication of rice beans.

6 Conclusion

A long-term, affordable solution to the problem of malnutrition is to incorporate dietary heterogeneity into the food system in order to attain zero hunger. The rice bean is a rich source of high-quality protein, soluble fiber, minerals, and antioxidants, as well as a broad range of health supplements and pharmaceuticals that can cure or prevent diseases influenced by food. Due to its several uses as manure, fodder, and food for people, it is advantageous and economically viable for farms in underdeveloped areas. As a result, there are large gaps in current research on rice beans and marketing, necessitating increased effort. Farmers should be encouraged to participate in research projects on underutilized local plant varieties like rice beans, and policies should be developed to support this. Scientists should investigate the cultivars that both farmers and consumers favor and undertake awareness campaigns to emphasize the advantages of the crop for malnourished people in underdeveloped countries in order to encourage its planting. The aim is to increase crop yield, remove toxic anti-nutritional components, shorten the time needed to cook, gain consumer acceptance, acquire market approval, and address the protein- and food-insecurity problem. It requires a lot of effort to bring these nutrient-rich crops at the top of any country's food chain. However, recent innovations like CRISPR/Cas9, speed breeding, genomic selection, and other high-throughput approaches could expedite the breeding process (Fig. 3).

Acknowledgements The authors thank the ICAR scheme – National Agricultural Science Fund (NASF/GTR – 7025/2018-19) for financially supporting this research. We also thank the Department of Biotechnology (DBT) Funds (BT/PR38339/GET/119/334/2020) for additional support towards completing this work.

References

Andersen P (2007) Food security through rice bean research in India and Nepal (FOSRIN). Report 3. In: Hollington PA (ed) Nutritional qualities of rice bean. Department of Geography, Universität Bergen/CAZS Natural Resources, College of Natural Sciences, Bangor University, Bergen/Bangor. Available from http://www.ricebean.org/publication.htm

Andersen P (2012) Challenges for under-utilized crops illustrated by rice bean (*Vigna umbellata*) in India and Nepal. Int J Agric Sustain 10:164–174

Andersen P, Kumar N, Acharya BP (2009) Rice bean food preparation and diets. https://hdl.handle.net/1956/6611

Arora RK, Chandel PS, Joshi BS, Pant KC (1980) Rice bean: the tribal pulse of Eastern India. Econ Bot 34:260–263

Bajaj M (2014) Nutrients and antinutrients in rice bean (*Vigna umbellata*) varieties as effected by soaking and pressure cooking. Asian Journal of Dairying & Foods Research 33(1):71–74. ISSN: 0976-0563. https://doi.org/10.5958/j.0976-0563.33.1.015

Bajracharya J, Singh S, Dangol B, Hollington PA, Witcombe JR (2008) Food security through rice bean research in India and Nepal (FOSRIN). Report 2. In: Identification of polymorphic markers. Agriculture Botany Division, Nepal Agriculture Research Council/CAZS Natural Resources, College of Natural Sciences, Bangor

Basavaprabhu NM, Niranjanmurthy NM, Asif M, Venkatesha KT, Vijay Kumar KV (2013) Genetic divergence analysis in rice bean, *Vigna umbellata* (L.). Intern. J Plant Sci 8(1):165–167

Basu P, Scholten B (2012) Technological and social dimensions of the green revolution: connecting pasts and futures. Int J Agric Sustain 10(2):109–116

Bepary RH, Wadikar DD, Neog SB, Patki PE (2017) Studies on physico-chemical and cooking characteristics of rice bean varieties grown in the NE region of India. J Food Sci Technol 54(4): 973–986. https://doi.org/10.1007/s13197-016-2400-z

Bhardwaj N, Kaur J, Sharma AP (2021) The beans and the peas from orphan to mainstream crops. Proc Qual Nutraceutical Appl 55–66. https://doi.org/10.1016/B978-0-12-821450-3.00012-3

Bhat R, Karim AA (2009) Exploring the nutritional potential of wild and underutilized legumes. Compr Rev Food Sci Food Saf 8:305–333. https://doi.org/10.1111/j.1541-4337.2009.00084.x

Bisht IS, Singh M (2013) Asian Vigna. In: Singh M, Upadhayay HD, Bisht IS (eds) Genetic and genomic resources of grain legume improvement. Elsevier, Oxford, pp 237–267

Bisht IS, Bhat KV, Lakhanpaul S, Latha M, Jayan PK, Biswas BK, Singh AK (2005) Diversity and genetic resources of wild *Vigna* species in India. Genet Resour Crop Evol 52:53–68

Buergelt D, von Oppen M, Yadavendra JP (2009) Quality parameters in relation to consumer's preferences in rice bean. In: Presentation at the International conference on grain legumes: quality improvement, value addition and trade, IIPR, Kanpur, India, February 14–16, 2009

Burkill IH (1953) The hooker lecture. Proc Linn Soc Lond 164:12. https://doi.org/10.1111/j.1095-8312.1953.tb00664.x

Campbell M, Holt C, Moore B, Yandell M (2014) Genome annotation and curation using MAKER and MAKER-P. Curr Proto Bioinformat 48:4.11.1–4.11.39

Chen H, Chen X, Tian J, Yang Y, Liu Z, Hao X, Wang L, Wang S, Liang J, Zhang L, Yin F, Cheng X (2016) Development of gene-based SSR markers in Rice bean (*Vigna umbellata* L.) based on transcriptome data. PLoS One 11(3):e0151040. https://doi.org/10.1371/journal.pone.0151040

Conti MV, Campanaro A, Coccetti P, De Giuseppe R, Galimberti A, Labra M et al (2019) Potential role of neglected and underutilized plant species in improving women's empowerment and nutrition in areas of subSaharan Africa. Nutr Rev 77:817–828. https://doi.org/10.1093/nutrit/nuz038

Dahipahle AV, Kumar S, Sharma N, Singh H, Kashyap S, Meena H (2017) Rice bean – a multipurpose, underutilized, potential nutritive fodder legume – a review. J Pure Appl Microbiol 11(1):433–439

Dhanasekar P, Reddy KS (2014) A novel mutation in *TFL1* homolog affecting determinacy in cowpea (*Vigna unguiculata*). Mol Genet Genomics 290:55–65. https://doi.org/10.1007/s00438-014-0899-0

Dhillon PK, Tanwar B (2018) Rice bean: a healthy and cost-effective alternative for crop and food diversity in: food security: the science, sociology and economics of food production and access to food. Int Soc Plant Pathol 10(3):525–535

Doi K, Kaga A, Tomooka N, Vaughan DA (2002) Molecular phylogeny of genus *Vigna* subgenus *Ceratotropis* based on rDNA ITS and atpB-rbcL intergenic spacer for cpDNA sequences. Genetica 114:129–145

Fan W, Lou HQ, Gong YL, Liu MY, Wang ZQ, Yang JL, Zheng SJ (2014) Identification of early al-responsive genes in rice bean (*Vigna umbellata*) roots provides new clues to molecular mechanisms of Al toxicity and tolerance. Plant Cell Environ 37:1586–1597

González AM, Yuste-Lisbona FJ, Saburido S, Bretones S, De Ron AM, Lozano R, Santalla M (2016) Major contribution of flowering time and vegetative growth to plant production in common bean as deduced from a comparative genetic mapping. Front Plant Sci 7:1940. https://doi.org/10.3389/fpls.2016.01940

Iangrai B, Pattanayak ADEA, Khongwi G, Pale EM, Gatphoh A, Das NK, Chrungoo (2017) Development and characterization of a new set of genomic microsatellite markers in rice bean (*Vigna umbellata* (Thunb.) Ohwi and Ohashi) and their utilization in genetic diversity analysis of collections from North East India. PLoS One 12(7):0179801

Isemura T, Kaga A, Tomooka N, Shimizu T, Vaughan DA (2010) The genetics of domestication of rice bean, *Vigna umbellata*. Ann Bot 106:927–944. https://doi.org/10.1093/aob/mcq188

Jayasundara S (2015) Harvest index of three-grain legume crops grown in Canada, calculated by two approaches. https://www.researchgate.net/post/How_is_harvest_indexestimated_especially_in_indeterminate_pulses

Joshi KD, Bhandari B, Gautam R, Bajracharya J, Hollington P (2007) Rice bean: a multipurpose underutilized legume. Paper presented at the 5th international symposium on new crops and uses their role in a rapidly changing world. The University of Southampton, Southampton, pp 3–4

Kang Y, Kim S, Kim M, Lestari P, Kim K, Ha B, Jun T, Hwang W, Lee T, Lee J, Shim S, Yoon M, Jang Y, Han K, Taeprayoon P, Yoon N, Somta P, Tanya P, Kim K, Gwag J, Moon J, Lee Y, Park B, Bombarely A, Doyle J, Jackson S, Schafleitner R, Srinives P, Varshney R, Lee S (2014) Genome sequence of mungbean and insights into evolution within *Vigna* species. Nat Commun 5(1):5443

Kang Y, Satyawan D, Shim S, Lee T, Lee J, Hwang W, Kim S, Lestari P, Laosatit K, Kim K, Ha T, Chitikineni A, Kim M, Ko J, Gwag J, Moon J, Lee Y, Park B, Varshney R, Lee S (2015) Draft genome sequence of adzuki bean, *Vigna angularis*. Sci Rep 5(1):8069

Kashiwaba K, Tomooka N, Kaga A, Han OK, Vaughan DA (2003) Characterization of resistance to three bruchid species (*Callosobruchus* spp., Coleoptera, Bruchidae) in cultivated rice bean (*Vigna umbellata*). J Econ Entomol 96:207–213

Katoch R (2013) Nutritional potential of rice bean (*Vigna umbellata*): an underutilized legume. J Food Sci 78:C8–C16

Katoch R (2020) Rice bean (Exploiting the Nutritional Potential of an Underutilized Legume). https://doi.org/10.1007/978-981-15-5293-9

Katoch R, Singh SK, Thakur N, Dutt S, Yadav SK, Shukle R (2014) Cloning, characterization, expression analysis and inhibition studies of a novel gene encoding Bowman-type protease inhibitor from rice bean. Gene 546:342–351

Kaul T, Eswaran M, Thangaraj A, Meyyazhagan A, Nehra M, Raman N, Bharti J, Gayacharan BC, Balamurali B (2019a) Rice bean (*Vigna umbellata*) draft genome sequence: unravelling the late flowering and unpalatability related genomic resources for efficient domestication of this underutilized crop. BioRxiv. https://doi.org/10.1101/816595

Kaul T, Raman NM, Eswaran M, Thangaraj A, Verma R, Sony SK, Sathelly KM, Kaul R, Yadava P, Agrawal PK (2019b) Data mining by pluralistic approach on CRISPR gene editing in plants. Front Plant Sci 10:801. https://doi.org/10.3389/fpls.2019.00801

Kaul T, Sony SK, Verma R, Motelb KFA, Prakash AT, Eswaran M, Bharti J, Nehra M, Kaul R (2020) Revisiting CRISPR/Cas-mediated crop improvement: special focus on nutrition. J Biosci 45:137

Kaul T, Easwaran M, Thangaraj A, Meyyazhagan A, Nehra M, Raman NM, Verma R, Sony SK, Abdel KF, Bharti J, Gayacharan, Badapanda C, Balasubramanian B. (2022) De novo genome assembly of rice bean (Vigna umbellata) – A nominated nutritionally rich future crop reveals novel insights into flowering potential, habit, and palatability centric – traits for efficient domestication. Front Plant Sci.13:739654. https://doi.org/10.3389/fpls.2022.739654.

Khadka K, Acharya BD (2009) Cultivation practices of rice bean. Local initiatives for biodiversity. Research and Development (LI-BIRD), Pokhara, p 28

Khan MM, Rafii MY, Ramlee IS, Jusoh M, Al-Mamun S (2021) Bambara groundnut (*Vigna subterranea* L. Verdi): a crop for the new millennium, its genetic diversity, and improvements to mitigate future food and nutritional challenges. Sustainability 15:1–27. https://doi.org/10.1038/s41598-021-938675

Khanal AR, Khadka K, Poudel I, Joshi KD, Hollington P (2009) Report on farmers' local knowledge associated with the production, utilization and diversity of rice beans (*Vigna umbellata*) in Nepal. In: The Rice bean network: farmers' indigenous knowledge of rice bean in Nepal (report N°4), EC. 6th FP, project no. 032055, FOSRIN (Food Security through Rice bean research in India and Nepal)

Ku HK, Ha SH (2020) Improving nutritional and functional quality by genome editing of crops: status and perspectives. Front Plant Sci 11:577313. https://doi.org/10.3389/fpls2020577313

Liu MY, Lou HQ, Chen WW (2018) Two citrate transporters coordinately regulate citrate secretion from rice bean root tips under aluminium stress. Plant Cell Environ 41:809–822. https://doi.org/10.1111/pce.13150

Mayes S, Massawe FJ, Alderson PG, Roberts JA, Azam-Ali SN, Hermann M (2012) The potential for underutilized crops to improve the security of food production. J Exp Bot 63(3):1075–1079. https://doi.org/10.1093/jxb/err396

Muñoz-Amatriaín M, Mirebrahim H, Xu P, Wanamaker S, Luo M, Alhakami H, Alpert M, Atokple I, Batieno B, Boukar O, Bozdag S, Cisse N, Drabo I, Ehlers J, Farmer A, Fatokun (2017) Genome resources for climate-resilient cowpea, an essential crop for food security. Plant J 89(5):1042–1054

Muthusamy S, Kanagarajan S, Ponnusamy S (2008) The efficiency of RAPD and ISSR markers system in accessing genetic variation of rice bean (*Vigna umbellata*) landraces. Electron J Biotechnol 11(3):32–41

Nnamani CV, Ajayi SA, Oselebe HO, Atkinson CJ, Igboabuchi AN, Ezigbo EC (2017) Sphenostylisstenocarpa (ex. A. Rich.) harms., a fading genetic resource in a changing climate: prerequisite for conservation and sustainability. Plan Theory 6:30. https://doi.org/10.3390/plants6030030

Noda A (1951) Studies on native (primitive) varieties of crop plant: observations on a climbing variety of adzuki bean native to San'in district, Japan. Jpn J Crop Sci 21:134–135

Padulosi S, Hoeschle-Zeledon I (2004) Underutilized plant species: what are they? LEISA Magazine, March issue, pp 5–6

Padulosi S, Hoeschle-Zeledon I (2008) Crops for the future: paths out of poverty. Strategic plan 2009–2013, Bioversity International regional Office for Asia, the Pacific and Oceania, Selangor, Malaysia, p 16

Pandiyan M, Ramamoorthi N, Ganesh SK, Jebraj S, Pagarajan P, Balasubramanian P (2008) Broadening the genetic base and introgression of MYMY resistance and yield improvement through unexplored genes from wild relatives in mungbean. Plant Mut Rep 2:33–38

Pandiyan M, Senthil N, Ramamoorthi N, Muthiah AR, Tomooka N, Duncan V, Jayaraj T (2010) Interspecific hybridization of *Vigna radiata* x 13 wild *Vigna* species for developing MYMV donor. Elect J Plant Breed 1:600–610

Parker TA, Sassoum L, Paul G (2021) Pod shattering in grain legumes: emerging genetic andenvironment-related patterns. Plant Cell 33(2):179–199

Parvathi S, Kumar VJF (2006) Value added products from rice bean (*Vigna umbellata*). J Food Sci Technol 43(2):190–193

Pattanayak A, Ingrai B, Khongwir DEA, Gatpoh EM, Das A, Chrungoo NK (2018) Diversity analysis of rice bean (*Vigna umbellata* (Thunb.) Ohwi and Ohashi)) collections from North Eastern India using morpho-agronomic traits. Sci Hortic 242:170–180

Pattanayak A, Roy S, Sood S, Iangrai B, Banerjee A, Gupta S, Joshi D (2019) Rice bean: a lesser known pulse with well-recognized potential. Planta 250(3):873–890

Popoola J, Ojuederie O, Omonhinmin C, Adegbite A (2020) Neglected and underutilized legume crops: improvement and future prospects. In: Shah F, Khan Z, Iqbal A, Turan M, Olgun M (eds) Recent advances in grain crops research. IntechOpen, London, pp 123–144. https://doi.org/10.5772/intechopen.87069

Priyadarshini A, Tiwari BK, Rajauria G (2021) Sustainable food production systems: the potential of pulses. In: Tiwari BK, Gowen A, McKenna B. Pulse Foods (Eds.) (Second Edition), Academic Press (Cambridge, Massachusetts, United States), pp 487–506, ISBN 9780128181843, https://doi.org/10.1016/B978-0-12-818184-3.00019-2

Pugalenthi M, Vadivel V, Gurumoorthi P, Janardhanan K (2004) Comparative nutritional evaluation of little-known legumes, *Tamarindus indica*, *Erythrina indica* and *Sesbania bispinosa*. Trop Subtrop Agroecol 4(3):107–123

Rajerison R (2006) *Vigna umbellata* (Thunb.) Ohwi and Ohashi. In: Brink M, Belay G (eds) Cereals and pulses. PROTA1, Wageningen

Rana J, Sood SG, Arun NK, Lal H (2014) Rice bean variety VRB3 (Him Shakti). Indian J Genet 74:268

Rejaul HB, Wadikar DD, Patki PE (2016) Rice bean: nutritional vibrant bean of Himalayan belt (north East India). Nutr Food Sci 46(3):412–431. https://doi.org/10.1108/NFS-08-2015-0097

Sadana B, Hira CK, Singla N, Grewal H (2006) Nutritional evaluation of rice bean (*Vigna umbellata*) strains. J Food Sci Technol 43(5):516–518

Saharan K, Khetarpaul N, Bishnoi S (2002) Variability in physicochemical properties and nutrient composition of newly released rice bean and faba bean cultivars. J Food Comp Analys 15(2): 159–167

Sarma BK, Singh M, Gupta HS, Singh G, Srivastava LS (1995) Studies in rice bean germplasm: research bulletin no. 34. ICAR Research Complex for NEH Region, Barapani, p 35

Schenke D, Cai D (2020) Applications of CRISPR/Cas to improve crop disease resistance: beyond inactivation of susceptibility factors. Science 23(9):101478. https://doi.org/10.1016/jisci2020101478

Seehalak W, Tomooka N, Waranyuwat A, Thipyapong P, Laosuwan P, Kaga A (2006) Geneticdiversity of the Vigna germplasm from Thailand and neighbouring regions revealed by AFLPanalysis. Genet Resour Crop Evol 53:1043–1059

Somta P, Kaga A, Tomooka N (2006) Development of an interspecific Vigna linkage map between *Vigna umbellata* (Thunb) Ohwi and Ohashi and Vnakashimae (Ohwi) Ohwi and Ohashi andits use in analysis of bruchid resistance and comparative genomics. Plant Breed 125:77–84

Sthapit B, Padulosi S, Mal B (2010) Role of on-farm/in situ conservation and underutilized crops in the wake of climate change. Indian J Plant Genet Resour 23(2):145–156

Sudha M, Anusuya P, Ganesh NM, Karthikeyan A, Nagarajan P, Raveendran M, Senthil N, Pandiyan M, Angappan K, Balasubramanian P (2013) Molecular studies on mungbean (*Vigna radiata* (L.) Wilczek) and rice bean (*Vigna umbellata* (Thunb.) interspecific hybridization for Mungbean yellow mosaic virus resistance and development of species-specific SCAR marker for rice bean. Arch Phytopathol Plant Protec 46:503–517

Tabassum J, Ahmad S, Hussain B, Mawia AM, Zeb A, Ju L (2021) Applications and potential of genome-editing systems in rice improvement: current and future perspectives. Agronomy 11: 1359. https://doi.org/10.3390/agronomy11071359

Takahashi Y, Iseki K, Kitazawa K, Muto C, Somta P, Irie K, Naito K, Tomooka N (2015) A homoploid hybrid between wild Vigna species found in a limestone karst. Front. Plant Sci 6:1050

Thakur S, Bhardwaj N, Chahota RK (2017) Evaluation of genetic diversity in rice bean [Vignaumbellata (Thunb.) Ohwi and Ohashi] germplasm using SSR markers. Elect J Plant Breed 8:674–679

Tharanathan RN, Mahadevamma S (2003) Grain legumes – a boon to human nutrition. Trends Food Sci Technol 14:507–518

Tian J, Isemura T, Kaga A, Vaighan DA, Tomooka N (2013) Genetic diversity of the rice bean (*Vigna umbellata*) as assessed by SSR markers. Genome 56:717–727

Tomooka N, Lairungreang C, Nakeeraks P, Egawa Y, Thavarasook C (1991) Mung bean and the genetic resources. The final report of the cooperative research work between Thailand and Japan submitted to the National Research Council of Thailand, March Tropical Research Center, Tsukuba, Japan

Tomooka N, Kashiwaba K, Vaughan D, Ishimoto M, Egawa Y (2000) The effectiveness of evaluating wild species, searching for sources of resistance to bruchid beetle in the genus *Vigna* subspecies *Caratotropis*. Euphytica 115:27–41

Tomooka N, Maxted N, Thavarasook C, Jayasuriya AHM (2002) Two new species, new species combinations and sectional designations in *Vigna* subgenus *Ceratotropis* (Piper) Verdcourt (Leguminosae, Phaseoleae). Kew Bull 57:613–624

Tripathi A, Iswarya V, Singh N, Rawson A (2021) Chemistry of pulses – micronutrients processing, Qual Nutraceut Appl 61–68. https://doi.org/10.1016/B978-0-12-818184-3.00004

Wang LX, Chen HL, Peng B, Wu JX, Wang SH, Blair MW, Cheng XZ (2015) The transferability and polymorphism of mung bean SSR markers in rice bean germplasm. Mol Breed 35:77. https://doi.org/10.1007/s11032-015-0280-y

Wang W, Simmonds J, Pan Q, Davidson D, He F, Battal A, Akhunova A, Trick HN, Uauy C, Akhunov E (2018) Gene editing and mutagenesis reveal inter-cultivar differences and additivity inthe contribution of TaGW2 homoeologues to grain size and weight in wheat. Theor Appl Genet 131:2463–2475

Wang M, Wang Z, Mao Y, Lu Y, Yang R, Tao X, Zhu JK (2019) Optimizing base editors for improved efficiency and expanded editing scope in rice. Plant Biotechnol J 17(9):1697–1699. https://doi.org/10.1111/pbi.13124

Wei Y, Yan J, Long F, Lu G (2015) *Vigna umbellata* (Thunb.) Ohwi et Ohashi or *Vigna angularis* (Willd.) Ohwi et Ohashi (Chixiaodou, Rice Bean). In: Liu Y, Wang Z, Zhang J (eds) Dietary Chinese herbs chemistry, pharmacology and clinical evidence. Springer, New York, pp 551–560

Yang JL, Zhang L, Li YY, You JF, Wu P, Zheng SJ (2006) Citrate transporters play a critical role in aluminium-stimulated citrate efflux in rice bean (*Vigna umbellata*) roots. Ann Bot 97:579–584

Zafar SA, Zaidi SS, Gaba Y, Singla-Pareek SL, Dhankher OP, Li X, Mansoor SP (2020) A engineering abiotic stress tolerance via CRISPR/Cas-mediated genome editing. J Exp Bot 71 (2):470–479. https://doi.org/10.1093/jxb/erz476

Zhang J, Zhang H, Botella JR, Zhu JK (2018) Generation of new glutinous rice by CRISPR/Cas9-targeted mutagenesis of the Waxy gene in elite rice varieties. J Integr Plant Biol 60:369–375. https://doi.org/10.1111/jipb12620

Part IV

Fruit Crops

Improvement of Nutraceutical Traits of Banana: New Breeding Techniques

Jaindra Nath Tripathi, Valentine Otang Ntui, Mathiazhagan Malarvizhi, Samwel Muiruri, Kundapura V. Ravishankar, and Leena Tripathi

Contents

J. N. Tripathi · V. O. Ntui · L. Tripathi (✉)
International Institute of Tropical Agriculture (IITA), Nairobi, Kenya
e-mail: l.tripathi@cgiar.org; J.tripathi@cgiar.org

M. Malarvizhi · K. V. Ravishankar
Division of Biotechnology, ICAR Indian Institute of Horticultural Research, Bengaluru, India
e-mail: ravishankar.kv@icar.gov.in

S. Muiruri
International Institute of Tropical Agriculture (IITA), Nairobi, Kenya

Department of Plant Sciences, Kenyatta University, Nairobi, Kenya

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_33

Abstract

Banana (*Musa* spp.) is an herbaceous, everlasting green monocotyledonous plant belonging to the family *Musaceae*. It is a major staple crop after rice, maize, wheat, potato and cassava, and it has a high potential to contribute to food and nutrition security. It is an excellent fruit full of micronutrients, especially vitamin A, iron, potassium, and magnesium, and is a source of energy for millions of inhabitants of tropical and subtropical regions. Despite these qualities, banana is still lacking in various essential nutrients. The conventional breeding program aims to increase the nutritional quality of banana. Still, the program is facing severe challenges due to the sterile seedless nature of banana and the narrow genetic diversity of several banana cultivars. Accessibility of well-annotated *Musa* genome sequences and established transformation and gene-editing platforms can contribute to developing banana with high dietary value. Furthermore, banana can produce edible vaccines, paving the way for future syringe-less vaccine development. This book chapter describes the various aspects of nutrition and the health-related importance of banana.

Keywords

Banana · Biofortification · Bioinformatic databases · Banana diversity · Conventional banana breeding · Genetic engineering · Gene editing · Nutraceutical

1 Introduction

Banana (*Musa* spp.) is one of humankind's first-ever cultivated crops. It is still a significant fruit of the tropics and subtropics and is reachable globally for sale and consumption. Portuguese explorers introduced it to the African mainland from the Guinea Coast of West Africa. The term "plantain" (for cooking banana) was derived from the word "plantano" of the Spaniards. The term "banana" includes all edible varieties consumed as ripe fruits or cooked food. Banana was first known to the Arabs from very early times, as it appears in "The Holy Koran" as the "tree of paradise" – which is akin to the "tree of knowledge" in Christianity. Banana is an

evergreen, long-lasting, monocotyledonous herb of the *Musaceae* family. It is an essential food security crop in the tropic and subtropic regions and popular fruit globally. It is grown on over 11 million hectares of fertile land in more than 150 countries and islands, with an annual production of approximately 156 million tons (Tripathi et al. 2022). More than 1000 cultivars and landraces of bananas have been recorded (Heslop-Harrison and Schwarzacher 2007). The crop is grown mostly by smallholder farmers as food security and cash crops due to the crop being available the whole year. Most of the banana are cultivated for local consumption, and only about 15–20% enter the world markets for income generation. Banana is also crucial for animal feed, and the leaves can be used as roof covers for mud houses in some villages. The center of origin of the banana is believed to be Southeast Asia, particularly in the forest of Indonesia, Malaysia, and the Philippines, and its domestic cultivation is considered to have been initiated primarily in these countries. It has since spread across south Asia, Africa, and central and southern America. It is also argued that there are two domestication centers, one in Southeast Asia and the other in New Guinea. There are very diverse varieties of banana still found in these regions. Originally, all the cultivated edible seedless and tasty banana originated from two diploid wild species of *Musa acuminata* (genome AA) and *Musa balbisiana* (genome BB) of the genus *Musa* (Tripathi et al. 2022). Most familiar parthenocarpic banana cultivars are triploid (AAA, AAB, and ABB genome). Diploid groups (AA, AB, and BB genome) and tetraploids (AAAA, AAAB, AABB, and ABBB genome) are rarely cultivated. Cooking banana is the main portion of food intake for the local populations in the tropics. The dessert banana has a global distribution and marketing. The nutritional value of banana makes them one of the most important fruit crops. In terms of economic worth, they are the fifth most important agricultural produce traded globally. Throughout history, *Musa* has provided humans with food, medicine, clothing, tools, shelter, paper, and handicrafts. It might be referred to as the "first fruit crop" because it was initially cultivated when hunting and gathering were still the main methods of obtaining food.

The major objective of the banana breeding program is to develop high-yielding banana cultivars with resistance to pests and diseases and focus on the plant architecture to reduce the size of the false stem known as pseudostem, early flowering and maturity, and good quality of the fruit. The greatest challenge for banana breeders is to find suitable mating partners as seed-producing, diploid banana with disease resistance and other beneficial traits and tetraploids with seed fertility for use in banana breeding. Most of the edible banana cultivars are sterile and seedless due to parthenocarpy. These sterile banana cultivars are not suitable for breeding programs. The essential agronomic features are more prevalent in diploid parents. Consequently, the primary banana breeding projects concentrated on creating enhanced diploid parents with fertile seeds that could be crossed. To complement the conventional breeding efforts, it is essential to combine molecular breeding technology, especially transgenic and gene editing, to bypass the reproductive barrier for improving banana cultivars enrich with high nutritional qualities and high yielding. To cultivate sufficient food to fulfill the growing population's increasing demand for food security, more resilient food production technology is required. To

achieve the full potential, advanced breeding techniques like plant genetic transformation and genome editing need to be explored to complement conventional breeding programs. This book chapter aims to provide information on the valuable phytochemicals available in bananas and offer insightful evidence regarding the fruits' possible health advantages and use as a natural food antioxidant.

2 Description of Nutritional Constituents

Banana is a fruit with high nutritional importance and has proven to provide cures for some health challenges, such as constipation and depression. Both unripe green and ripe yellow banana contain 2.6 g dietary fiber, 75 g water, 23 g carbohydrates, 12 g sugars, 1 g proteins, a minimal amount of fat, and vitamins and minerals, especially Vitamin A, choline, B6, vitamin C, and potassium, phosphorus and magnesium, sodium, zinc, and several antioxidants mainly dopamine and catechin (fdc.nal.usda.gov). Yellow- and orange-pulped banana are rich in trans-β-carotene (Fungo and Pillay 2013). Consumption of carotenoid-rich fruits reduces various diseases, such as cardiovascular problems, type II diabetes, and cancer, and boosts immunity (Fungo and Pillay 2013). Specific banana cultivars are rich in provitamin A, an essential alternative for the poor population of the world who have been affected by severe vitamin A deficiency (Fungo and Pillay 2013). The tender core of banana pseudo-stem flour has higher flavonoids, polyphenols, insoluble dietary fiber, total dietary fiber, lignin, hemicellulose, cellulose, antioxidant capacity, and free-radical scavenging (Fungo and Pillay 2013). The male flower buds treat bronchitis, ulcers, and dysentery.

Cooked flowers are regarded as a healthy diabetic diet. Additionally, banana do not contain trace amounts of fat, sodium, or cholesterol, making them a healthy food option even for restrictive diet plans. Banana is considered a functional food due to their immunity-boosting properties. Fully Ripe banana peels comprise a very high number of phytochemicals known as antioxidants. They have been applied in traditional medicine in rural areas as an antiseptic to promote wound healing from insect bites and burns and tans human history (Pereira and Maraschin 2015). Eating a ripe banana helps to relieve constipation; similarly, eating an unripe banana provides a cure for loose motion. Banana is also a source of vitamin C, essential for preventing skin diseases in small children. The antioxidants in banana fruits, like superoxide ions, hydrogen peroxide, hydroxyl radicals, nitric oxide radicals, and singlet oxygen, have been linked to the treatment of a variety of diseases, including diabetes, age-related muscular degeneration, arthritis, some cancers, genotoxicity, inflammation, and Alzheimer's disease (Septembre-Malaterre et al. 2016). The tender core of banana pseudo-stem is used for food and medicinal properties in South India and Northeastern states due to its high phenolic content (Kandasamy and Aradhya 2014). The extract from the core of the stem is thought to be effective in reducing the size and weight of kidney and bladder stones. Banana peels are a rich source of antioxidant phytochemicals, such as delphinidin, anthocyanins, and cyanidins.

Banana is an excellent source of potassium. A single banana has around 23% of the daily recommended potassium intake. Additionally, current research indicates that potassium helps lower blood pressure in people who are deficient in it. Potassium also lowers the chance of having a stroke. Bananas have long been known for their antacid properties, which help to prevent stomach ulcers. Leucocyanidin, a banana flavonoid, significantly thickens the stomach's mucous membrane layer. Banana consumption is an excellent method to treat heartburn since they assist in neutralizing acidity.

Banana is a good source of natural antioxidants for foods. Mainly gallocatechin gallate is more copious in peel (158 mg/100 g wt.) than in pulp (29.6 mg/100 g (Someya et al. 2002). Similarly, banana contains another antioxidant called dopamine, chemically known as catecholamine, both in pulp and peels. Peels contain more dopamine (80–550 mg/100 g) than pulp (2.5–10 mg/100) (Kanazawa and Sakakibara 2000). The bioactive compounds, a product of secondary metabolism, have a beneficial potential by contributing to antioxidant activities. Carotenoids, flavonoids, and biogenic amines are the primary phytochemicals in banana fruits and are well-known to enhance human health.

Banana fruit ripening is a complex biochemical pathway with various activities to improve the fruits' aroma, nutritional quality, color, and texture. Green unripe mature fruits contain many carbon sources in the form of starch (12–35%), which gradually converts into sucrose (80%), glucose (10%), and fructose (10%) as a soluble sugar at the time of late ripening. Understanding the mechanisms that regulate starch conversion into soluble sugars during ripening is essential to minimizing postharvest losses by preventing fast ripening. Ethylene and other hormones are responsible for fruit ripening in banana which converts starch to soluble sugars (Cordenunsi-Lysenko et al. 2019). At the molecular level, it was understood that MaMADS24 and MaMADS49 proteins interact with several *MaMADS* genes for banana ripening, ethylene-signaling genes, and starch solubilization genes (Liu et al. 2017). The promoter regions of several genes also subsidize the mechanism by which ethylene regulates starch accumulation in ripened banana. About 200 AP2/to devise the strategy to prevent postharvest losses due to ripening in the natural conditions hindering the transportation for export purposes.

3 Genetic Resources of Health-Related (HR) Genes

Thousands of banana cultivars have evolved from natural crosses of the diploid A genome of *M. acuminata* Colla and B genome of *M. balbisiana* Colla. At Ethylene Responsive Factor Binding Protein (EREBP), transcription factors (TF) were known in the *Musa* genome (D'Hont et al. 2012). The effects of ethylene hormones and TF on the conversion of the starch-to-sugar pathway in banana remain a challenging trait to understand, and it needs further biotechnological interventions. The banana cultivars include diploids, triploids, and tetraploid genomic constitutions that are primarily sterile and vegetatively propagated. Based on consumption, they are generally classified as

"banana," which are sweet and eaten uncooked as fruit; "cooking banana," such as the East African Highland Banana (EAHB) that is cooked while unripe, "plantains" that are starchy and roasted while unripe, and "beer banana" are used for fermentation and deep frying (Heslop-Harrison and Schwarzacher 2007). The plantain subgroup of the banana belongs to AAB genomic group, while the dessert banana belongs to the AA, AB, AAA, and AAB genomic groups, and the cooking banana falls under AAA genomic group. The hybridization between the diploid species (A and B) resulted in common edible triploid cultivars and tetraploids. The commercially critical cultivated banana belongs to the group AAA. It includes 'Gros Michael', 'Cavendish', 'Grand Naine', 'Robusta', etc. The banana cultivars are classified according to their genome composition (Table 1).

The Cavendish banana is the preferred dessert banana due to its high TSS and vitamin C levels, particularly in 'High gate', 'Gros Michael', 'Jari buaya', and

Table 1 Banana cultivars based on genome constitution. (Adapted from Nayar (2010)

Genomic composition	Banana cultivars
AA	'Inarnibal', 'Paka', 'Matti', 'Anakomban', 'Pisang Jari Buaya', 'Pisang Lilin', 'Senorita', 'Kadali', 'Sucrier', 'Kluai Khai', 'Lady's Finger', 'Orito', 'Pisang Mas'
AAA	'Ambon', 'Cavendish' ('Dwarf Cavendish', 'Giant Cavendish', 'Grand Nain', 'Williams'), 'Gros Michel' ('Cocos', 'Highgate', 'Lowgate'), 'Ibola', 'Basrai', 'Lujugira-Mutika', 'Pisang Masak Hijau' (Lacatan), 'Red' ('Green Red'), 'Robusta' ('Harichal', 'Malbhog'), East African Highland Banana ('Musakala', 'Nakabulu', 'Nakitembe', 'Nfunka', 'Mbidde')
AAAA	'Pisang Ustrali'
BB	'Bhimkol', 'Biguihan', 'Gubao', 'Pa-a Dalaga', 'Tani'
BBB	'Abuhon', 'Inabaniko', 'Lap Chang Kut', 'Mundo', 'Saba Sa Hapon', 'Saba', 'Sabang Poti', 'Turrangkog'
AB	'Kunnan' ('Adukkan', 'Poonkalli', 'Poovilla Chundan'), 'Ney Poovan' ('Kisubi', 'Safed Velchi'), 'Sukali Ndizi' ('Kumarangasenge')
AAB	'False horn' ('French', 'French horn'), 'Laknau', 'Maia Maoli', 'Moongil', 'Mysore' ('Sugandhi'), 'Nendran', 'Pisang Raja', 'Plantain Horn', 'Pome' ('Pachanadan', 'Pacovan', 'Prata Ana', 'Virupakshi'), 'Popoulu', 'Ilohena', 'Rasthali', 'Silk'
ABB	'Bluggoe' ('Nalla Bontha', 'Pisang Batu', 'Punda'), 'Pisang Awak' ('Klue Namwa', 'Karpuravalli', 'Pey Kunnan', 'Yawa'), 'Monthan', 'Peyan', 'Klue Teparot', 'Pelipita', 'Kalapua', 'Cardaba'
AAAB	'Atan'
AABB	'Kalamagol', 'Laknau Der'
ABBB	'Bhat Manohar'
AS	'Aso', 'Kokor', 'Ungota', 'Vunamami'
AT	'Umbubu'
AAT	'Kabulupusa', 'Karoina', 'Mayalopa', 'Sar'
ABBT	'Giant Kalapur', 'Yawa 2'
Unknown	'Fei'

A Musa acuminata, *B Musa balbisiana*, *S Musa schizocarpa*, *T Musa textiles*

'Bucaneiro' have high total phenols and total carotenoid content among cultivars with AAA and AAAA genome (Borges et al. 2014). Studies have shown that cultivars from Papua New Guinea (Englberger et al. 2010) and Malaysia (Borges et al. 2014) possess high amounts of provitamin A. The orange and yellow-fleshed cultivars contained high β-carotene than the cream or white-fleshed cultivars, indicating a correlation between orange/yellow pulp color and carotenoid content. Further, Red Dacca, Latundan, and Cavendish banana varieties were reported to have inhibitory activities against α-amylase and α-glucosidase enzymes and high free radical scavenging abilities, which make them suitable for consumption by type 2 diabetes patients (Adedayo et al. 2016). In general, sugars, minerals, proteins, polyphenols, and carotenoid contents increased during ripening, whereas fiber, lipids, and starch contents were higher in unripe fruits (Kookal and Thimmaiah 2018). Interestingly, banana pseudostem and peel are excellent sources of antioxidants and polyphenols. The pseudostem of a few banana cultivars have high antioxidant properties due to its high phenol and flavonoid content. Besides, various banana and plantain varieties belonging to *M. acuminata*, *M. balbisiana*, and *M. paradisiaca* have shown antimicrobial activity against bacterial and fungal pathogens (Mostafa 2021).

Some banana cultivars that are rich in nutrients and metabolites and valuable for human health are listed (Table 2). Many cultivars were reported to be rich in multiple nutrients like phenolic compounds, minerals, vitamins, TSS, and starch. For example, the cultivar 'Nendran' is high in carotenoids, phenolics, minerals, vitamins, and total soluble sugars. Similarly, cultivar 'Saba', 'Highgate', 'Njali Poovan', 'Karpooravalli', and 'Grand Naine' contains multiple health-related nutrients. Such cultivars can serve as valuable genetic resources for molecular mapping and breeding.

4 Classical Genetics and Traditional Breeding

Conventional breeding is difficult in *Musa* species due to the sterility and high heterozygosity. Hence, developing large segregating mapping populations of triploid, cultivated banana cultivars that are female and male-sterile is cumbersome. Moreover, basic genetic and inheritance studies in banana were hindered in the past due to constant pressure on breeders' need to develop disease and pest-resistant varieties as the banana-growing areas were continuously threatened by conditions such as fusarium wilt and Sigatoka. A few studies had previously examined the inheritance pattern of genes for vegetative and fruit traits using mapping populations or accessions derived from diploid fertile cultivars. Over the years, researchers have attempted to determine the ancestry of the A and B genomes in banana which contributed to the formation of cultivars with different genetic backgrounds worldwide. It was established that the maternal transmission of chloroplast genome and paternal information of mitochondrial genome in banana using Restriction Fragment Length Polymorphism (RFLP) markers. Haplotyping of mitochondrial and chloroplast genomes has supported the theory of the A and B genomes have a common ancestor. In addition, the study established the presence of six chloroplasts and seven

Table 2 Some banana cultivars with high health related traits

Properties	Cultivar name	Reference
Antimicrobial activity	'Dole', 'Fougamou', 'Kluai Tiparot', 'Mbwazirume', 'Namwah Khom', 'Petit Naine', 'Pelipita', and 'Saba'	Jouneghani et al. (2020)
High carotenoid content	'Aibwo/Suria', 'Asupina', 'Berlin', 'Gatagata/Vudito', 'Fagufagu', 'Horn Plantain', 'Hongjiaowang', 'Jari Buaya', 'Jaran', 'Karpooravalli', 'Kirkirnan', 'Kluai Khai Bonng', 'Ropa', 'Wasolay', 'Highgate', 'Thap Maeo', 'Malbut, 'Lahi', 'Apantu', 'Bira', 'To'o', 'Sepi', 'HungTu', 'Parao', 'Pisang Raja', 'Pacific Plantain', 'Wain', 'Red Dacca', 'Lakatan', 'Sucrier', 'Utin Iap', 'Karat', 'Utimwas', 'Iemwahn', 'Utiak', 'Dukerehda', 'Red Banana', 'Nendran', 'Goldfinger' and 'Hua Moa', 'Toraka', 'Warowaro'	Englberger et al. (2010), Ekesa et al. (2015), and Kargar (2019)
High vitamin content	'Karat', 'Aibwo/Suria', 'Gatagata/Vudito', 'Fagufagu', and 'Akeakesusu'	Englberger et al. (2010)
High phenolic content	'Jaran', 'Wasolay', 'Nam', 'Highgate', 'Gros Michel', 'Thap Maeo', 'Saba', 'Champa Madras', 'Bucaneiro', 'Ney Poovan', 'Tiparot', 'Njali poovan', 'Nendran', 'Ducasse', 'Ladyfinger', 'Terrinha', 'Marmelo', 'Ouro', 'Red Banana', 'Poovan', 'Muomva-red', 'Grande Naine', 'Rasbale', 'Pisang Raja', 'Goldfinger', and 'Kandarian'	Bashmil et al. (2021) and Khoza et al. (2021)
High mineral content	'Khai', 'Ouro da Mata', and 'Pacha Nadam', 'Akondro mainty', 'Makyughu 1', 'Ijihu inkundu', 'NIGAB-1', 'NIGAB-2', 'Rasakadali', 'Nendran', 'Saba', 'FHIA-01', 'Basrai', 'Jawari', 'South Tenerife', 'North Tenerife'	Khoza et al. (2021)
High protein content	'Makyughu 2', 'Muraru', 'Akondro mainty', 'Finger Rose', 'Pisang Awak'	Khoza et al. (2021)
High folic acid content	'Lampung'	Ningsih and Megia (2019)
High TSS content	'Kadali', 'Rasakadali', and 'Nendran', 'Njali poovan', 'Robusta', 'Ouro', 'Nanica', 'Nanicão', 'Caru-Roxa', 'Prata', 'Prata-Anã', 'Mysore', 'Williams', and 'Pacovan'	Kookal and Thimmaiah (2018)
High vitamin C/antioxidant content	'Nendran', 'Njali poovan', 'Ducasse', 'Ouro', 'Terrinha', 'Muomva-red', 'Luvhele', and 'Grande Naine'	Bashmil et al. (2021)
High starch content	'Marmelo', 'Terrinha', 'Karpooravalli', 'FHIA-01', 'Macho', 'Enano', and 'Valery'	Khoza et al. (2021)

mitochondrial cytoplasmic-gene pools among the *M. acuminata* and *M. balbisiana* accessions (Boonruangrod et al. 2008). Cytogenetic studies through karyotyping have led to an understanding of chromosomal structural arrangements of genotypes with different genomes in banana. These results indicate that the banana genome is

constituted by common ancestors who were evolved by *M. balbisiana* (BB genome) and *M. acuminata* (AA genome) (Dehery et al. 2021).

Interestingly, the invention of next-generation sequencing technology validated the cytogenetic finding by identifying two structural variations in A and B genomes, thereby influencing local recombination and causing a segregation distortion and aneuploidy in edible interspecific triploid hybrids (Baurens et al. 2019). About inheritance of fruit quality traits in cultivar 'Sukali Ndizi', TSS exhibited nonadditive and dominant gene action. In contrast, pulp texture and flavor traits were found to be influenced by additive and complementary gene action. Fruit acidity showed an incomplete dominance effect, with a single major gene and multiple genes controlling the trait (Buregyeya et al. 2021). Furthermore, phenotypic characteristics like finger weight and bunch weight were shown to strongly correlate positively with yield-related attributes in banana. Albinism in banana is a harmful trait caused by a lack of chlorophyll in the plant tissues. Using full-sib plantain-banana hybrids, researchers determined that the segregating ratio of albinism was 15:1, which might be due to the complementary gene action of one or two recessive alleles. Dwarfism is a desirable commercial trait in banana due to the toppling of large-size pseudostem after bearing the big bunches of the fruits. Dwarfism in segregating population generated by crossing triploid cultivar 'Bobby Tannap' (AAB) and diploid 'Calcutta 4' manifested a trisomic test-cross segregation ratio of 2:1, indicating that the trait was regulated by a single recessive gene, known as a dwarf gene. Ortiz (1997) reported the presence and inheritance of 2n pollen in *Musa* diploid and polyploidy cultivars. A single dominant gene that controls 2n pollen facilitates the introgression of beneficial alleles from diploid species to polyploidy cultivars through sexual polyploidization. However, banana lacks extensive studies on the inheritance patterns of fruit traits. An increased focus on improving cultivars with favorable fruit characters and disease and pest resistance can be achieved only when genetic mechanisms underlying these traits are thoroughly known. Therefore, adapting molecular breeding approaches will ease the burden of conventional breeding methods in developing seedless, vegetatively propagated banana progenies.

5 Genetic Diversity Analysis of Banana

Knowledge of the genetic diversity of banana germplasm, cultivars, and wild types is crucial for understanding the genetic makeup and their suitability in breeding programs. Lots of genetic diversity is observed in the plantain and the desert banana. Different morphometric indicators and molecular marker-based diversity studies have been reported in banana over the years. The phenotypic polymorphisms among plantain cultivars based on yield-related traits can serve as indicators for the field performance of cultivars (Brisibe et al. 2021). The important phenotypic traits like size and weight of bunch per plant, the total number of fruits per bunch, the girth of the pseudostem, height of the plant, time period from flowering to harvest, the color of the leaf, rachis type, peduncle hairiness, fruit shape, fruit apex, etc., were used to study the diversity among banana cultivars and hybrids (Brisibe et al. 2021). Similarly, descriptors like petiole canal, bract curling, and color have helped in differentiating cultivars with different ploidy levels as well as between cooking and dessert banana. In

addition, mineral profiling of fresh fruit pulp has been used as an indicator to study variability among Indian *Musa* accessions (Devarajan et al. 2021).

Molecular marker-based diversity analysis aid in minimizing discrepancies in phenotyping and distinguishing cultivars with high phenotypic similarities. Molecular markers have facilitated studies on interspecific and intraspecific diversity, relatedness, and population structure among banana cultivars. It includes random amplified polymorphic DNA (RAPD; Wahyudi et al. 2020), Inter Simple Sequence Repeat (ISSR; Wahyudi et al. 2020), Simple-sequence repeats (SSR; Brisibe and Ubi 2020), Internal transcribed spacer (ITS), Sequence-related amplified polymorphism (SRAP), *MAT*K marker (Hariyanto et al. 2021), Conserved DNA-Derived Polymorphism, Start Codon Targeted (SCoT; Igwe et al. 2022), Directed Amplification of Minisatellite-region DNA (DAMD; Pinar et al. 2019), and Inter-Retrotransposon Amplified Polymorphism (IRAP) markers (Saraswathi et al. 2020). Microsatellite markers are employed to analyze the population structure and genetic diversity of banana (Mertens et al. 2021). Besides, rbcL plastid gene sequences were utilized to understand the genetic relatedness and diversity among the cultivars and wild species of *M. acuminata* (AA genome), *M. balbisiana* (BB genome), and *M. paradisiaca* (AAB, ABB genome). This rbcL sequence-based barcoding can be useful in species identification and discrimination in banana (Ainiyah et al. 2020). Biswas et al. (2020) recently reported the development and validation of functionally relevant genic SSR markers in *Musa* species which were mapped to *M. acuminata*, *M. balbisiana*, and *M. Schizocarpa* genomes. Similarly, genic SSR markers were developed from the transcriptome of cultivars 'Bee Hee Kela' and 'Calcutta-4' (Venkataramana et al. 2015). These SSR markers can be used in genetic diversity studies, germplasm characterization, population structure analysis, and comparative mapping not only in banana but also in related monocot species. In addition, methylation-sensitive Amplified Fragment Length Polymorphism (AFLP) markers were used to determine the epigenetic variations among the genetically identical East African Highland banana. These heritable epigenetic changes in the offspring can lead to different functional traits. Consequently, using epigenetic markers in diversity analysis can aid in the conservation of epigenetic resources and their integration in breeding programs Further, genotyping of the external transcribed spacer region of the ribosomal DNA locus of *M. acuminata* and *M. balbisiana* subspecies led to the development of a SCAR marker that can differentiate the gene pools and identify the wild ancestors (Jeensae et al. 2021). Phenotyping and genotyping employing morphological descriptors and molecular markers, respectively, will be an efficient method for comprehending the variability and relatedness of *Musa* species.

6 Molecular Analysis of Hypersensitive Response Genes and Quantitative Trait Loci on Banana Genome

Mapping population studies for the banana genome are faced with various constraints like parthenocarpy, low fertility, polyploidy, heterozygosity, and small population size. In a molecular mapping study, the first requirement is generating

mapping populations. Diploids, fertile cultivars, wild types, and triploids have been used to create mapping populations to examine the segregation pattern of genes involved in different quantitative traits and disease resistance (Ahmad et al. 2020). Very few efforts were made to generate a mapping population and identify QTLs concerning biotic and abiotic resistance in *Musa* species. Thus, there are very few studies regarding QTLs or genes related to fruit or other morphological traits in banana. Recently, Biobiany et al. (2022) have used genotyping by sequencing to identify QTLs linked to fruit traits like pulp acidity, firmness, and dry matter content. Twelve prominent QTLs were identified, among which a QTL for pulp acidity was located in the LG1_7 of the genetic map of 'Pisang Madhu'.

Further, a di-haploid mitochondrial genome of *M. acuminate subsp. malaccensis* DH-Pahang was sequenced and assembled. Similar efforts will help improve our understanding of cytoplasmic male sterility and flower and seed development in banana, as these traits are controlled by the genes present in the mitochondrial DNA (Prakash et al. 2022). Despite the lack of mapping studies related to fruit traits, the availability of complete genome sequences of *M. acuminata* and *M. balbisiana* has enabled genome-wide identification of several transcription factors (Wang et al. 2022) and flavonoid gene families (Pandey et al. 2016) that are involved in plant development, fruit ripening, various metabolism and plant stress responses from biotic and abiotic stresses. The genes identified in these studies could help assist banana breeding program initiatives in improving the fruit quality of the existing cultivars.

7 Genomics-Aided Breeding for HR Traits

The seedless trait (parthenocarpy combined with female sterility) was studied utilizing Genome-wide association studies using DArt markers in 105 diploid accessions with *M. acuminata* genetic background. This helped identify six candidate genomic regions with two genes related to female sterility (Sardos et al. 2016). Interestingly, Genome-wide association studies (GWAS) on 33 accessions from Papua New Guinea confirmed the six-candidate regions and discovered an orthologous gene to Histidine Kinase CKI1 that may control the female sterility trait (Sardos et al. 2016). Similarly, GWAS and genomic selection (GS) using prediction-based models were reported in a multiploidy training population of 307 banana genotypes. Out of the six prediction models employed, the BayesB model had the highest prediction capability for fruit-filling and fruit bunch traits.

Further, more models considering additive gene effects performed better in prediction Also, multiple QTLs with significant results are located on chromosome 3, responsible for controlling the fruit-filling trait in banana (Nyine et al. 2019). Further, transcriptome analysis has shed light on the expression pattern of genes involved in carotenoid metabolism in the 'Xiangfen1' banana (Dong et al. 2022). In the future, "Omics" based studies can help identify genes controlling critical health-related traits, supporting genomics-assisted breeding efforts. Henceforth, the use of

GWAS and genomic selection in banana breeding provides new possibilities for improving selection efficiency, particularly for fruit quality attributes.

8 Recent Concepts and Strategies for Banana Improvement

The global population is increasing day by day at an alarming rate. By 2050, the population of the world is expected to surpass 9 billion. This increasing population burden on the planet and the serious impact of climate change on agriculture production will fuel a massive demand for food, and it is estimated that nearly 1 billion people suffer from nutritional and food insecurity. Increased production of food crops with high nutritional quality using advanced technologies such as genome editing promises to address these insecurities. Plant genome editing tools are a group of advanced genetic engineering and molecular biology techniques that allow the plant biotechnologist to knock out faulty genes with high precision. Currently, several gene editing technologies are available in the hand of plant biotechnologists for crop improvement especially site-directed nucleases (SDNs), zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and the clustered regularly interspaced short palindromic repeats/CRISPR-associated protein (CRISPR/Cas). These techniques have been applied to manipulate plant genes in several crop species for increased nutritional quality. These gene editing tools cause specific double-strand breaks (DSBs) in DNA at targeted sites to modify genes in a highly precise method with minimal or no off-type effects. The double-stranded break in the DNA is repaired by the self-healing mechanism in the plant cells, either of nonhomologous end joining (NHEJ) or homology-directed repair (HDR) if there is a donor template available, leading to small insertions, deletions, substitutions of nucleotides, or gene replacement depending on the cell cycle phase and the presence or absence of a repair template containing homologous terminal sequences (Fig. 1). There are three types of editing techniques especially site-directed nucleases 1, 2 and 3 (SDN1, SDN2, and SDN3), have been used with respect to the DNA repair mechanism (Tripathi et al. 2020). SDN1 is effective and is based on NHEJ error-prone repair of DBS in the host genome, resulting in gene silencing, gene knockout, or alteration in the gene function. SDN2 is less efficient and has high reliability and is achieved when a repair template that is like the DSB is added to the CRISPR machinery. The DSB is then repaired via HDR, producing nucleotide substitution or targeted indels. In SDN3, a donor template that is longer than the homologous sequences in which the DSB was formed is used to repair the DSB via HDR. Based on the donor sequence, SDN3-type repair in DNA results in the insertion of the full gene or part of genetic material being used at the target location. The repair is less efficient and reliable.

8.1 Zinc Finger Nucleases (ZFN)

Zinc finger nucleases (ZFN) are synthetic endonucleases that include zinc finger DNA-binding region and a nuclease subunit, similar to the type of II restriction enzyme FokI endonuclease (Fig. 1). ZFN contains C2H2 zinc-finger domains, with each finger recognizing three nucleotides on the specific gene target. ZFNs have

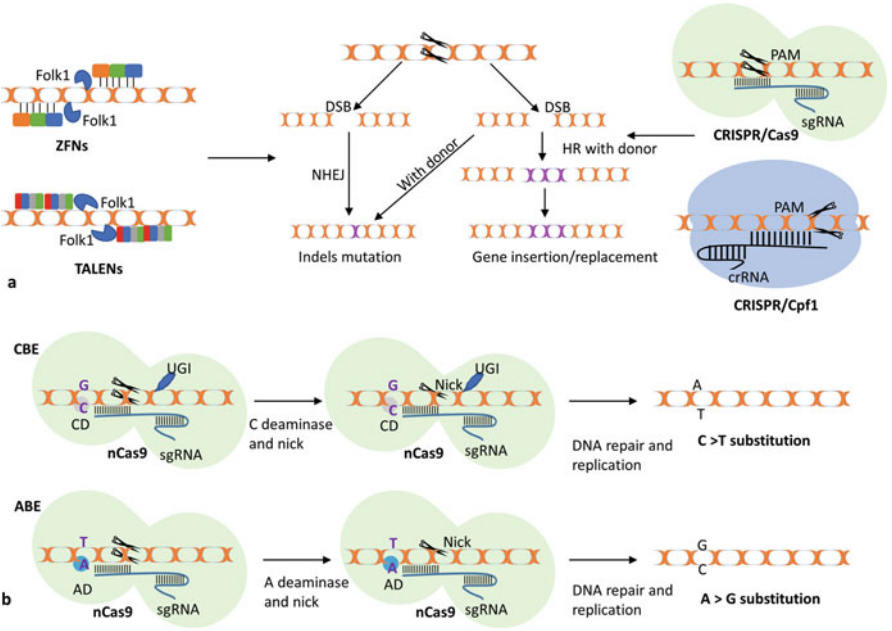


Fig. 1 Commonly used genome editing tools used for genetic manipulation of plants. (**a**) Genome editing tools which create double-stranded break (DSB). The break, shown in the middle, is restored either by homology recombination (HR) or nonhomologous end-joining (NHEJ). NHEJ inserts small indels (insertion, deletion, or substitution) into the DSB site of the genome and results in frame-shift mutations or premature stop codons. HR can cause small indels, gene substitution, and insertions depending on the presence or absence of a homologous donor DNA. (**a**) Zinc finger nucleases (ZFNs) and Transcription activator-like effector nucleases (TALEN) on the left panel use FokI endonuclease to cut DNA double strands. Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated protein 9 system on the right panel employs single-guide RNA (sgRNA) for DNA binding and Cas9 protein for DNA cleavage. While CRISPR/Cpf1 system uses mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) for DNA binding and Cpf1 protein for DNA cleavage. CRISPR/Cas9 recognizes a G-rich protospacer adjacent motif (PAM) while CRISPR/Cpf1 recognizes a T-rich PAM. (**b**) Diagrammatic presentation of CRISPR/Cas9-mediated base editing technique. In the cytosine base editor (CBE) system, cytosine (C) in the targeting region replaces with uracil (U), then U is replaced with thymine (T) in DNA repair or replication processes, creating a C-G to T-A substitution. In the Adenine base editor (ABE) technique, adenine (A) in the genome targeting region is converted to inosine (I), which is used as guanine (G) by polymerases, creating A-T to G-C substitutions. Abbreviations used in the diagram: *ZFN* - zinc-finger nuclease, *TALEN* - transcription activator-like effector nuclease, *DSB* - double-strand break, *NHEJ* - nonhomologous end joining, *HR* - homologous recombination, *PAM* - protospacer adjacent motif, *CRISPR* - clustered regularly interspaced short palindromic repeats, *Cas9* - CRISPR associated protein 9, *sgRNA* - single-guide RNA, *crRNA* - CRISPR RNA, *ABE* - adenine deaminases-mediated base editing, *CD* - cytidine deaminases, *AD* - adenine deaminases, *CBE* - cytidine deaminase-mediated base editing, *nCas9* - Cas9 nickase, *UGI* - uracil glycosylase inhibitor

successfully been used to target endogenous genes in Arabidopsis, tobacco, and soybean, creating desired mutations, gene replacements, deletions, or inversions. The main issues with ZFNs, however, are their limited specificity, which causes severe off-type effects in the plant genome, and their excessive complexity in the

designing of the construct, which causes a lower rate of the success of the endonuclease enzymes (Tripathi et al. 2020). These limitations of ZFN led to the discovery of the upgraded version of the endonucleases known as *transcription activator-like effector nucleases* (TALEN).

8.2 Transcription Activator-Like Effector Proteins (TALEN)

TALENs are synthetic restriction enzymes that contain FokI endonuclease subunit with a DNA-binding site and repeats of the DNA restriction domain to cut any desired site of the genome. TALEN proteins are secreted by Xanthomonas bacteria linked with the Fok1 nuclease domain. It is similar to ZFN proteins rely on the FokI nuclease subunit functioning as a dimer, with two monomers designed for each genomic target (Fig. 1). However, unlike ZFNs, TALENs possess a broader targeting range and are less challenging to engineer, appear to be more mutagenic than ZFNs, and are highly specific. TALEN-directed mutagenesis has been used in rice to target resistance against bacterial blight disease (Tripathi et al. 2020), and many plant species have been successfully edited utilizing this technology.

With the availability of banana genomic sequences, both ZFN and TALEN can be used for genome editing in a banana to increase nutritional quality. Unfortunately, these technologies require expertise in molecular cloning as well as knowledge of DNA binding sites for individually targeting applications. Recently, CRISPR/Cas9 has empowered researchers to perform genome editing in plants more conveniently, precisely, and safely with minimal off-type effects.

8.3 Clustered Regularly Interspaced Palindromic Repeats (CRISPR)/Cas9

The CRISPR/Cas9 system is based on small RNA-guided which bind on the targeted site of the genome and Cas proteins that cut the specific site of DNA to provide a double-stranded break, in comparison to TALENs and ZFNs, which are based on protein-DNA binding endonuclease. Theoretically, CRISPR/Cas9 technology originated from bacterial type II CRISPR/Cas immune system, which protects them against an invasion of bacteriophages and viruses. Because of its simplicity, design flexibility, high efficacy, and ability to simultaneously edit several genes, CRISPR/Cas9 technology has become the most popular and widely used technique (Ntui et al. 2020).

CRISPR/Cas9 system has two main components, the small guide RNA (gRNA) and the Cas proteins (Fig. 1). The Cas proteins are endonuclease that recognizes target DNA by matching the 5′ leading sequences of gRNA with the 5′ leading sequences of DNA. It also recognizes the adjacent protospacer motif (PAM) sequence and starts editing the DNA upstream. The gRNA consists of a framework of a 20-nucleotide user-defined spacer sequence that directs the Cas endonuclease to a target sequence in the specific genome that is complementary to the 20 nucleotides before the adjacent protospacer motif (PAM). The PAM is a three-nucleotide

sequence, typically NGG or NAG, found near the cleavage site of the target DNA. The main function of the Cas proteins is to start DSB in the DNA upstream in the targeted region of the genome (Ntui et al. 2020).

8.4 CRISPR12a and Cas13a Systems

There are several other Cas proteins have been discovered and applied for plant genome editing, for example, Cas12a (Cpfl) and Cas13a. Cas12a (Cpf1) is a class 2 type V-CRISPR that consists of RuvC sites that belong to the retroviral family. It comprises crRNA biogenesis RNase and single-stranded DNase activity and functions by first chopping up target DNA and processing its crRNA. It was developed from Prevotella and Francisella 1 bacteria (Roy and Soni 2021). It recognizes T-rich PAM, TTN/TTTN/TTTV (N = A/T/C/G; V = A/C/G) which is situated upstream of the target site, hence cutting DNA at the distal location of the PAM. It uses RuvC endonuclease, which is regulated by a single RNA. The endonuclease then recognizes PAM, which is rich thymidine-rich and produces staggering cuts (Roy and Soni 2021. Cas12a has shown more efficient multiplex gene manipulation by a single sequence array on the selected gRNA (Roy and Soni 2021). Cas13a is a class 2 type VI-A ribonuclease capable of targeting and cleaving the phage genome, which consists of single-stranded RNA molecules. It has shown potential for use in detecting RNA viruses, with more accuracy and specificity than PCR detection. A PAM region is not required for Cas13a ribonuclease to detect the RNA viruses in plants (Aman et al. 2018).

8.5 Base Editing and Prime Editing System

The Cas proteins use NHEJ and HDR processes to repair DSB in the target genome. These approaches have disadvantages, including decreased delivery, efficiency, and unwanted rearrangements caused by DSB at the targeted locus. To refine the gene editing with minimal damage to the parent genome, an upgraded version of CRISPR/Cas9 editing known as base editing technique (BE) was invented. BE is used for target manipulation of a single base pair without involving DSB in the genome. The base-editing tools use Cas9 nickase (nCas9) or inactive Cas9 fused to an endonuclease enzyme with single base pair conversion activity. In the base editing technique, a guide RNA binds to the target DNA, and when the bases pair, a DNA bubble having a short segment of ssDNA is produced (Roy and Soni 2021). A DNA deaminase fused to an inactive Cas9 or (dCas9) enables the base substitution at single-nucleotide resolution. Base-editor tools having cytosine deaminases convert cytosine (C) to thymine (T), creating a C-G to T-A or A-T to G-C substitution.

Similarly, the adenine deaminase base editor converts adenine (A) to guanine (G), making A-T to G-C substitutions (Fig. 1). The DNA deaminases function as effectors allowing C-G to T-A or substitution. At the same time, the RNA-guided CRISPR system acts as a genomic locator of the targeted locus. Cytidine-deaminase

base editing (CBE) has been used in several economically important crops. Zong et al. 2018). Adenine-deaminase base editor (ABE) is more difficult to use than CBE. However, researchers developed Adenine-deaminase effectors, which they can use to generate point mutations in some crops (Kang et al. 2018).

Prime editing is another tool developed to achieve gene editing in various organisms. It works the same way as traditional CRISPR/Cas systems, allowing DNA base pair replacements, minor insertions, and deletions (indels) (Chen et al. 2021). Primer editing, on the other hand, does not cause DSB and does not necessitate the use of a foreign DNA segment; it addresses frameshifts caused by indels and reduces off-target effects. To edit the genome, a fusion protein composed of Cas9 H840A nickase connected to a designed RT enzyme and long gRNA, known as pegRNA, is required. Although prime editing has the potential to supplement current CRISPR editing methods and enables precise and targeted DNA alterations, the biological process has yet to be well known, even though prime editing can be a fascinating tool to modify the genome of a banana.

8.6 CRISPR Activation

CRISPR/Cas9 gene editing tools have supported the creation of new technologies that rectify some of the limitations of traditional genetic editing systems and revolutionize many fields of plant science. The CRISPR/Cas9 transcriptional activator technique (CRISPRa) embraces a lot of potential to generate improved crops with desirable agronomic traits. CRISPRa is a form of CRISPR tool that induces gene expression by combining transcriptional activators with a modified version of Cas9 devoid of endonuclease activity (dead Cas proteins, also known as dCas). When the nuclease activity of the Cas9 protein is converted into a deactivated form (dCas9), CRISPR/dCas9 retains the ability to bind the targeted DNA sequence but does not cause DSB on the targeted site of the genome. This combination of dCas9 and transcriptional activation sites enables the precise and effective transcriptional activation of any genes without causing DSB at the DNA of targeted sequences of endogenous genes.

The most common type of CRISPR activator is VP64, a tetramer of VP16, a well-characterized transcription activator from the herpes simplex virus. VP64 has been shown to increase endogenous gene expression. It was one of the first CRISPRa systems developed, and it has demonstrated a significant induction gene over-expression in many crops (Shakirova et al. 2020). CRISPRa systems of the first generation have two components: dCas9 coupled with transcription factors and sgRNA. P65 and p300, two more first-generation dCas trans-activators, were also synthesized and employed for gene activation (as activators, various oligomers of VP16 have been developed).

According to Lowder et al. (2018), a dCas9 containingVP64 system including the deactivated CRISPR-associated protein 9 (dCas9) coupled with four tandem repetitions of the transcriptional activator VP16 (VP64) was shown to activate endogenous genes in plants. Arabidopsis and tobacco observed effective transcriptional

activation of protein-coding and noncoding genes. However, the gene activation rate of the first-generation of CRISPRa system with single-domain Cas9 fusions showed only low/moderate overexpression of the targeted endogenous gene in plants. The Second-generation CRISPRa system has also been explored.

The second-generation system consists of dCas9, sgRNA, and effectors recruited in multiple copies by specific dCas9 or sgRNA domains. This type of arrangement can cause changes in gene expression, be it activation, repression, epigenetic alterations, or something else (Shakirova et al. 2020). These systems include the following: (1) the scaffold and casilo, both of which are based on the scaffold RNA (sRNA). (2) The Synergistic Activation Mediator, based on a chimeric MS2-p65-HSF1 activation helper protein and chimeric dCas9-VP64, sgRNA with synthetic aptamers for MS2 recruitment, and a chimeric MS2-p65-HSF1 activation helper protein. The Supernova Tagging System is effective because antibodies have a high affinity and specificity for short peptide sequences. Lowder et al. (2018) used the second generation to create a CRISPRa system that targeted protein-coding and noncoding genes previously manipulated in Arabidopsis by the dCas9-VP64 system. They used a modified guide RNA scaffold gRNA2.0 (designated CRISPR-Act2.0) with VP64 fused to dCas9 and discovered that the CRISPR-Act2.0 system activated transcription more efficiently than the dCas9-VP64 system alone.

Selma et al. (2019) have developed a CRISPRa system in *Nicotiana benthamiana* using dCasEV2.1 loaded with a six-gRNA combination targeting the promoters of the NbDRF and NbAN2 genes. Compared to the first-generation CRISPRa system, transgenic lines harboring the second-generation CRISPRa system demonstrated significant gene activation (Selma et al. 2019). Pan et al. (2021) recently established a highly robust CRISPR-Acct3.0 system for inducing several genes in rice, tomato, and Arabidopsis. The CRISPR–Act3.0 system was developed by experimenting with several transcription activators based on deactivated Streptococcus pyogenes Cas9 (dSpCas9). Compared to the CRISPR-Act2.0 and the first-generation systems, the CRISPR–Act3.0 system activated genes four to six times more (Pan et al. 2021).

In our laboratory at IITA-Kenya, we are using CRISPRa to induce the overexpression of endogenous banana genes, including antibacterial Vicilin, Leucine-Rich Repeat, Wall Associated Kinases, Pathogenesis-Related gene, and disease resistance R gene by transcriptional factor inducers to confer resistance to banana Xanthomonas wilt. The endogenous *Musa* genes were annotated based on transcriptome analysis of bacterial disease-resistant progenitor *Musa balbisiana* and disease-susceptible Pisang Awak (Tripathi et al. 2020). Six constructs were designed and delivered to embryogenic cells, each targeting the promoters of three endogenous genes. Initial screening of the regenerated plants revealed targeted gene overexpression ranging from two- to eightfold by relative real-time qPCR, and further, these events evaluated in the screen house showed enhanced resistance against *Xanthomonas campestris* pv. *musacearum*. Apart from overexpressing disease resistance genes, the CRISPRa system can also activate genes to increase the nutritional composition of banana cultivars.

8.7 Gene Editing of Banana

A couple of studies have been documented on gene editing in banana. Currently, the genome editing strategy focuses on the CRISPR/Cas9 knockdown of endogenous genes. It is now possible to precisely target and alter banana genes thanks to the availability of reference genome sequences. The first gene editing was reported in the banana cultivar 'Rasthali' containing the AAB genome using phytoene desaturase (PDS) as a visual marker gene (Kaur et al. 2020). The researchers utilized a single gRNA to knockout in the PDS gene, producing albino phenotypes in the generated plants. However, the mutation efficiency was detected at about 59%. In the same year, Naim et al. (2018) reported employing polycistronic gRNAs to edit the PDS gene in 'Cavendish Williams' with the AAA genome, and mutation efficiency was detected at about 100%.

Ntui et al. (2020) also reported the generation of edited events with two gRNAs targeting the PDS gene to achieve 100% mutation efficiency in sweet banana 'Sukali Ndiizi' (AAB genome) and plantain 'Gonja Manjaya' (AAB genome). The PDS is a visual marker frequently used to optimize gene editing protocol in plants. Phytoene desaturase is an enzyme in the primary carotenoid pathway that alters phytoene into carotenoid precursor's phytofluene and ζ-carotene. When its function is knocked out, plants with albino, variegated, or pale green phenotypes are generated, and they can be seen with the naked eye, depending on the mutation pattern.

Generating disease-resistant plants with increased economic value, food, and nutritional enhancement are the main objectives of the genome editing of the banana. The major banana diseases that affect yield are the banana streak virus (BSV) and banana Xanthomonas wilt (BXW). BSV is a dsDNA badnavirus that integrates into the B genome of plantain (AAB genome). CRISPR/Cas9-based editing was performed to inactivate the endogenous banana streak virus (eBSV) in the plantain cultivar 'Gonja Manjaya', rendering functional viral proteins which were unable to develop new coat proteins (Tripathi et al. 2020). Further, it has been investigated that CRISPR/Cas9-mediated genome editing of Musa Downy mildew resistance 6 (Musa DMR6) in desert banana 'Sukali Ndiizi' led to enhanced resistance to bacterial disease (Tripathi et al. 2022).

Apart from disease resistance, genome editing has also been investigated to increase the shelf life of banana fruits and change plant architecture. In the banana cultivar 'Gros Michel,' CRISPR/Cas9 technology was applied to generate disruption in the gibberellin 20ox2 (MaGA20ox2) gene of M. acuminata. The gibberellin pathway was disrupted, resulting in shorter plants (Shao et al. 2020). CRISPR/Cas9 was exploited by Hu et al. (2021) to alter aminocyclopropane-1-carboxylase oxidase (MaACO1) in M. acuminata (AAA group, cv. Brazilian). ACO is a significant enzyme of the ethylene biosynthetic pathway, which uses a reduction process to convert ACC to ethylene. The MaACO1 gene was disrupted, resulting in plants that grew and developed similarly to control plants but were significantly shorter. Furthermore, the fruits of the MaACO1 mutants were somewhat shorter, lighter, and had a 60-day delay in ripening compared to the control (Hu et al. 2021).

8.8 Genome Editing of Banana for Nutritional Quality

Increasing specific vitamins and essential components in food crops to increase their nutritional content is gaining considerable attention. This practice is known as biofortification. It is a cost-effective method of increasing vitamin and mineral content in food crops, hence alleviating malnutrition (hidden hunger). Cereals, roots, tubers, and banana are vital staple crops in various parts of the world. These crops may be high in carbs, but they may be lacking in quality proteins and micronutrients. Malnutrition and its related disorders will be reduced by increasing protein and micronutrient content in these crops.

The application of gene editing technology has been used in several crops, especially camelina, grape, potato, rapeseed, rice, sweet potato, tomato, and wheat, to enhance nutritional qualities (Liu et al. 2021). Only one study has been published using a CRISPR/Cas9-based strategy to biofortify banana. By targeting to edit lycopene epsilon-cyclase (LEC) gene with CRISPR/Cas9 technology, Kaur et al. (2020) increased β-carotene concentration in the 'Grand Naine'. They reported that the edited events exhibited a sixfold higher concentration of β-carotene in the fruit pulp than the nonedited plants. These strategies can increase bananas' nutritional contents, especially amino acids, iron, zinc, and other nutrients.

8.9 External DNA Free Gene-Editing

Banana genome editing for nutritional quality using CRISPR/Cas9 and plasmid delivery might be accomplished with high efficiency. On the other hand, the genome-edited banana obtained through plasmid delivery contains selection marker genes and the Cas9 protein, which are introduced into plant cells by *Agrobacterium*. Because breeding to remove these selection marker genes and Cas protein sequences is impossible in asexually propagated bananas, the mutated plants may be classified as GMOs by regulatory authorities in some countries, decreasing public acceptance. It is necessary to develop gene-edited banana without any external DNA that will evade strict scrutiny and be acceptable to the regulators and consumers without any unwanted fear and assumptions. To create DNA-free genome-modified banana plants, various protocols could be used. A protocol was developed to insert Cas9 protein-gRNA ribonucleoproteins (RNPs) complexes into plant cells (Liang et al. 2017). After mutation, these RGENs-RNPs complexes dissolve themselves and are removed by the cellular mechanism in the plant cells, minimizing off-target effects and leaving no evidence of external DNA sequences in the plant genome (Liang et al. 2017). There are various methods to deliver RGENs-RNPs into plant cells by application of electroporation, particle delivery system, protoplast transfection, coated silica nanoparticle, and polyethylene glycol method, using cell-penetrating proteins. The generation of plants from banana protoplasts remains difficult and time-consuming work, even though protoplasts are excellent starting material to produce external DNA sequences free plants. Even though plant regeneration from protoplasts, isolated from embryogenic cells from banana, has been reported by

several researchers in the past. Still, most of the protocols need to be reproducible for plant regeneration. However, the protoplasts' high fragility, meager plant regeneration efficiency, nonreproducibility of protocols, complex regeneration, and a high percentage of somaclonal changes make this technique more challenging. Researchers are devoting more time and effort to solving the bottleneck of the protoplast regeneration system to speed the development of external DNA-free gene-edited plants in banana and several other crops.

Another strategy is to express the editing compounds into the plant cell transiently. *Agrobacterium* infection can be applied to deliver the editing complexes transiently and produce external sequences-free plants without the plasmid DNA being integrated into the plant genome. This technology has been proven in tobacco, and 8.2% of nontransgenic plants were generated by momentary expression of CRISPR/Cas9-encoded gRNAs, knocking out the *PDS* gene (Chen et al. 2018). Veillet et al. (2019) used a similar technique, and they developed a cytidine-based editor to mutate the acetolactate synthase (ALS) gene in potatoes and tomatoes via *Agrobacterium* transient infection (CBE). They successfully generated potato and tomato plants with high mutation efficiency. However, several undesired effects were seen in the generated plants, indicating that the approach needs to be tweaked further. The critical bottleneck of both RNP complexes and momentary expression is that, without a visible marker, identifying possible mutants is difficult due to the thousands of generated plants that must be evaluated to select the mutants. The most effective way to insert CRISPR/Cas9 components into plant cells is by *an Agrobacterium*-mediated system with appropriate antibiotic selection. However, plasmids with T-DNA excision and removal features after editing are required. Costa et al. (2020) proposed two strategies for removing T-DNA from mutants. The primary method is based on the Flippase (Flp) site-specific recombinase, which distinguishes the 34 bp long FRT sequences (Flp/FRT) system, and another technique is based on Cas9 and synthetic cleavage target sites (CTS) near T-DNA boundaries, which are recognized by the sgRNA. Unluckily, adornment at T-DNA boundaries was discovered, damaging the t-DNA removal mechanisms. Despite this, the research represents a significant step forward in synthesizing DNA-free CRISPR plants using *Agrobacterium* harboring plasmid and antibiotic selection.

We are vigorously developing a methodology for producing external DNA-free banana using the momentary *Agrobacterium* system of the Cas9-gRNA reagent targeting the *PDS* gene in our laboratory. Additionally, we are optimizing various protocols for protoplast isolation, PEG transfection, and plant regeneration. If this happened, it would open up the robust technique to produce external DNA-free plants with increased nutritional contents that will be unfettered, pass through present GMO legislation, and be acceptable to consumers and producers globally.

8.10 Nanotechnology

Nanotechnology deals with nanoparticles to prevent postharvest losses of fruits and vegetables. The mostly edible nanoparticles of 1–100 nm coat fruits sold at distant

locations. Edible coating of nanoparticles like chitosan, zinc, and silver is a potential technology for delaying the ripening and increasing the self-life of banana fruits because nanoparticle coatings are simple and affordable. Cold chains are not required for storage. Nanoparticles also have biopolymer properties, which protect fruits from bacterial, fungal, and other pathogens (Odetayo et al. 2022).

9 Genetic Modification of Banana

Biofortification of banana with provitamin A and iron are the primary health-related trait for genetic modification. The production of plant-based edible vaccines is a primary focus for researchers in future programs for human diseases. The low-income families are highly suffering from micronutrient deficiencies due to insufficient fulfillment of daily dietary requirements. Vitamin A deficiency is one of the critical health complications in children and lactating mothers globally, especially in developing countries, even though massive food fortification and supplement programs have been implemented. Using genetic engineering, biofortification of banana is a potential strategy to develop provitamin A (PVA) enriched banana. Paul et al. (2017) reported overexpression of the carotenoid biosynthetic pathway enzyme phytoene synthase as a strategy to increase PVA and prolong fruit maturation time in banana fruit. Iron is another micronutrient essential for human health, mostly in growing children and pregnant and lactating mothers. Yadav et al. (2017) reported enhancing the iron in the 'Rasthali' cultivar of banana (AAB) using *MusaFer*1. Similarly, Kumar et al. (2011) enhanced iron and zinc contents in 'Rasthali' with soybean ferritin cDNA using two different plasmid constructs, pSF, and pEFE-SF.

Banana as a fruit can be an excellent choice for developing edible vaccines due to the availability of an optimized transformation and regeneration system and the fact that the fruits are eaten raw. Edible plant-based vaccines do not require cold chains and high-end manufacturing facilities that reduce the cost of vaccination. Edible vaccines can be a realistic approach if researchers from medicine, biotechnology, and agronomy work together; this will eliminate the pains caused by the syringe in infants.

Genetic improvement of banana commenced during the 1990s to develop a protocol of clonally propagated plants by protoplast electroporation, microprojectile bombardment, and *Agrobacterium*-mediated gene insertion. Embryogenic cells (EC) are favored for genetic transformation in banana due to high regeneration and transformation efficacy and reduced chimerism. EC regeneration systems and transformation protocols have been optimized for several banana and plantain cultivars. However, the production of embryogenic cells is a laborious, lengthy, and extremely cultivar-dependent process (Tripathi et al. 2022). These protocols have been established to develop high-yielding disease and pest resistance climate-smart banana cultivars (Tripathi et al. 2022). An efficient *Agrobacterium*-mediated genetic transformation platform for banana and plantain is fully operational at IITA-Kenya (Tripathi et al. 2022). Our laboratory can generate hundreds of transgenic and gene-edited events annually for several cultivars of plantain and banana.

In the absence of natural resistance in the cultivated edible banana to *Xanthomonas campestris* pv. *musacearum* and the various bottleneck of conventional breeding favor biotechnological applications like genetic engineering and gene editing. Transgenic bananas overexpressing the *Capsicum annuum* hypersensitive response-assisting protein or plant ferredoxin-like protein genes were generated and evaluated in the greenhouse and as well as in a confined field in Uganda (Tripathi et al. 2020). These transgenic bananas showed resistance to *Xanthomonas campestris pv. musacearum* (Xcm) and good agronomic performance comparable to control uninfected plants (Tripathi et al. 2020). Two or more genes can be stacked in a construct and delivered into banana cells. In the cells, these genes will act independently to produce the desired trait. Gene stacking has been achieved in banana for disease resistance. For example, transgenic banana developed using stacked gene constructs *Hrap* and *Pflp* showed enhanced resistance against bacterial pathogen Xcm (Tripathi et al. 2022). Besides, reduced disease symptoms of *Fusarium oxysporum* f.sp. *cubense* was seen in the two events of banana cultivar Rasthali (AAB) transformed by stacked antimicrobial plasmid construct (*Ace-AMP1* and *Pflp*) (Sunisha et al. 2020). Plant-parasitic nematodes incur huge losses on plantains and cooking bananas. Transgenic plantain transformed with antifeedant cysteine proteinase inhibitor and synthetic peptide-related genes showed enhanced resistance against *Radopholus similis* and *Helicotylenchus multicinctus* in the confined field trial in Uganda (Tripathi et al. 2022).

RNA interference (RNAi) induced gene silencing approach is a posttranscriptional gene silencing mechanism applied to study functional genomics of a particular gene and related products with high efficiency, specificity, and easiness in several crops. RNAi-induced technology has been proven in several other crops, especially maize, rice, wheat, and Arabidopsis. RNAi-induced gene silencing of the *gusA* gene was reported in banana, where the regenerated plants showed a reduction in gene expression and enzyme activity (Dang et al. 2014). RNAi was used to control the banana bunchy top virus (Elayabalan et al. 2013). Further, RNAi technology was applied to control banana aphids (*Pentalonia nigronervosa)* in transgenic banana (Cavendish Williams) and plantain (Gonja Manjaya and Orishele), targeting the silencing of the *AChE* gene (Jekayinoluwa et al. 2021).

Plant biotechnologists are exploring ways to bypass tight biosafety regulatory guidelines and public perceptions against genetically modified crops and their products by inserting plant genes from sexually compatible wild plants of the same species to develop new plant varieties with beneficial traits. All the genes, including the selection marker gene, promoter, and terminator, are derived from the same species. The cisgenesis technique is a potential tool for developing plant varieties that propagate vegetatively, such as potatoes, apples, and banana (Telem et al. 2013). It can directly improve an existing variety without disturbing the plant's genetic makeup by transferring beneficial trait-specific genes. The production of selection marker-free plants usually requires optimization of protocols to generate cisgenic plants. Kleidon et al. (2020) devised a technology for developing cisgenic banana plants through steroid-induced recombination that produced marker-free transgene events with minimal copy number.

10 Bioinformatics as a Tool

The first banana draft genome was obtained by genotype *Musa acuminata* subspecies *malaccensis* doubled-haploid genotype Pahang (DH-Pahang), yielding a 523 megabase genome. Shortly after this first sequencing, *M. balbisiana* variety 'Pisang Klutuk Wulung' (PKW) a diploid, was sequenced (Davey et al. 2013). Currently, the full-genome sequence of banana is accessible through the Banana genome hub platform (http://banana-genome.cirad.fr/) for the application of comparative genomics. In addition, bioinformaticians have created the 472 Mb sequence assembly of DH-Pahang and *Musa balbisiana* (Droc et al. 2013). Such comparative genomic approaches are essential for predicting functional gene annotations within *Musa* genomes.

Additional genomes of cultivated banana, their wild relatives, and those of closely related species have been sequenced to improve the database (Wang et al. 2019). The exponential increase in sequenced genomes has proportionately occurred in other plant species. The large sequencing projects have increased the demand for databases and tools to manage, process, and derive meaning from the vast amounts of data. This section will discuss the different databases and platforms currently available to handle various aspects of the processing of genome sequences. Characterization of an organism's genome involves functionally comparing the species to characterize the content and structure of its genes. This ultimately results in a deeper understanding of gene arrangements within the chromosomes, observations of any similarities or differences in the evolution, and identification of the functional significance of distinct genetic representations.

10.1 International Nucleotide Sequence Database Collaboration (INSDC)

Genomic information databases started in 1979 as the Los Alamos Sequence Database initiative to store biological sequences. This database was renamed GenBank 3 years later and relocated to the National Center for Biotechnology Information (NCBI), where it currently resides (Sayers et al. 2022). The creation of the International Nucleotide Sequence Database Collaboration (INSDC; http://www.insdc.org/), a joint initiative of the NCBI, the European Molecular Biology Laboratory (EMBL) and the DNA Databank of Japan (DDBJ) ushered in core infrastructure for sharing nucleotide sequence data and metadata to the public (Arita et al. 2021). The three INSDC nodes synchronize data daily to ensure its similarity. The collaboration is not limited to nucleotide data but also a collection of information on research projects deposited into the database of Bio Project.

10.2 Gene and Genome Databases of the NCBI

There are 13 NCBI databases with information about genes and genomes. Eight out of these 13 store genome and genome-associated data, while five are for gene storage (Sayers et al. 2022).

10.3 NCBI Genome Data Storage Databases

10.3.1 Nucleotide Database

This database collects nucleotide sequences from the GenBank, RefSeq, Third Party Annotation, and protein databases. The Nucleotide database is, in fact, a collection of different databases. The GenBank, for example, the National Institutes of Health (NIH) in the USA, genetic sequence databases, is an assortment of well-annotated gene sequences. The RefSeq is another component of the Nucleotide database and contains sequences of curated genomic DNAs, RNAs, and proteins (Sayers et al. 2022). RefSeqs stably reference genome annotation, gene identities and characteristics, analysis of mutations and polymorphisms, and expression and comparatives studies.

10.3.2 BioSample Database

The BioSample database stores descriptive information, aka metadata supplied by the submitter, and relating to biological materials from which the primary data archives are derived (Sayers et al. 2022). The NCBI's libraries have data from a wide range of samples from any species making the BioSample database equally diverse.

10.3.3 Sequence Read Archive (SRA) Database

The SRA comprises all sequencing data generated on the sequencing platforms, including Roche 454 GS System®, Illumina, Life Technologies AB SOLID System®, Helicos Biosciences Heliscope®, Complete Genomics®, Pacific Biosciences SMRT® and, Oxford nanopore® (Sayers et al. 2022).

10.3.4 Taxonomy Database

This database contains the names and phylogenies of more than 160,000 organisms containing molecular data within the NCBI. Continuous additions to the taxonomy database aimed at synchronizing it to other databases are an ongoing effort (Sayers et al. 2022).

10.3.5 The Assembly Database

This database contains information on the structural organization of genomes of various organisms, assembly names and related meta-data, reports on statistics, and links of genomic sequences (Sayers et al. 2022).

10.3.6 Bio Project Database

This database was also known as ASD Genome Project. It comprises data on genomics, including functional genomics, genetics, and the resulting dataset links. This database describes the project scope, material, and objectives. A provision for retrieval of datasets is challenging to find due to inconsistent annotation (Sayers et al. 2022).

10.3.7 Genome Database

This database is a collection of the complete and fully drafted genomes of the various organisms and related data, such as well annotated sequences, maps, chromosomes, assemblies (Sayers et al. 2022).

10.3.8 Bio Collections Database

This database stores curated culture collection metadata, museums, herbaria, and other collections of natural history. On display in the records group are codes, collections' institution information, and links to corresponding NCBI data (Sayers et al. 2022).

10.4 NCBI Gene Expression and Storage Databases

10.4.1 Gene Expression Omnibus (GEO) Profiles

This database stores individual gene expression and molecular abundance profiles assembled from the Gene Expression Omnibus (GEO) repository (Sayers et al. 2022).

10.4.2 Gene Database

The Gene database integrates information from many species, including nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and their corresponding genome-, phenotype-, and locus-links (Sayers et al. 2022).

10.4.3 GEO Data Sets Database

This database stores a curation of assembled gene expression and molecular abundance datasets from the GEO repository. The database can be queried to locate experiments of interest by entering search terms. Additional DataSet records include cluster tools and differential expression queries (Sayers et al. 2022).

10.4.4 PopSet Database

This database is a compilation of related DNA sequences from the population, phylogenetic, mutation, and ecosystem deposited with the GenBank. A population set contains information on genetic variation within an organism, whereas a phylogenetic group has sequences and their respective alignments for a single gene from several closely related organisms (Sayers et al. 2022).

10.4.5 HomoloGene Database

This database is a system for constructing putative homology groups from the complete gene sets of eukaryotic species and is automated (Sayers et al. 2022).

10.5 Comparative Plant Genome Databases

Comparative genomics involves analyzing genomes emphasizing similarities and differences at the sequence or the annotation level. The major types of comparison are pairwise involving two genomes or multiple whole-genome alignments involving more than two sequences. A wide array of comparative genomics databases has been established, primarily plant specific. The central banana-specific database is the Banana Genome Hub (Droc et al. 2013) (https://banana-genome-hub.southgreen.fr/).

The Banana Genome Hub is a public banana-specific hub providing genomic information in *Musa*. The strategy used in implementing Banana Genome Hub was exploiting interconnected generic software and establishing a reliable framework for those interested in banana research (Droc et al. 2013). The Banana Genome Hub depends on a robust comparative genomics database called GreenPhylDB, which contains the *Musa* protein-coding genes. GreenPhylDB includes protein gene families based on the automatic clustering of 22 whole plant genome sequences.

Various other plants comparative genomics databases that can serve the banana research community exist. One such is EnsemblPlants a (Bolser et al. 2016) **(**https://plants.ensembl.org/) EnsemblPlants is an integrative database carrying genome-wide information for sequenced plant species. This database provides data on the genome sequence, gene models, functional annotation and polymorphisms, the structure of populations, linkage, and phenotypic data (Bolser et al. 2016). Another nonbanana plant comparative genomic database important for the banana community is Phytozome (https://phytozome-next.jgi.doe.gov/). Phytozome was first released in 2008 to provide a centralized hub for use by people with varying degrees of computational sophistication. The data to be accessed in phytochrome include annotated plant gene families, the evolutionary history of gene families and individual genes, plant genes in their genomic context, and the assignment of putative function to uncharacterized user sequences, among others. The numbers and sophistication of existing comparative genomics platforms will continue growing owing to the elaborate sequencing efforts facilitated by advancements in high-throughput genome sequencing approaches.

10.6 Protein or Metabolome Databases

The major protein-related databases include (1) the NCBI Protein database, which comprises a collection of sequences from several sources (https://www.ncbi.nlm.nih.gov/). Uniprot is a protein database geared toward providing the scientific community with comprehensive, high-quality protein sequences, and functional information (UniProt Consortium 2021). The Uniprot provides well-annotated protein sequence data linked to summaries of experimentally verified or computationally predicted functional information about proteins (UniProt Consortium 2021).

10.7 Metabolome Databases

The most comprehensive metabolite and pathway database are the KEGG which comprises PATHWAY, BRITE, and MODULE databases aimed at understanding cellular and organismal processes (Kanehisa et al. 2017). The other metabolic pathway-specific database that features banana is the MetaCyc, where a total of 3006 pathways are described. The number of enzymes and transporters deposited in the MetaCyc is 17.1%, comprising 6128 enzymes, 112 transporters, and 36,528 proteins.

11 Political and Regulatory Issues

Significant concerns about genetically modified or gene-edited crops are common, especially public acceptance, loss of biodiversity, food allergy, the risk to natural environments, antibiotic resistance, creation of super weeds, ownership by big companies, and several mythical issues. Researchers have only been concerned about potential governmental support by funding and scientific evidence-based decision-making policies to avoid unnecessary implementation delays and reap the benefits of advanced scientific tools for food production to feed the growing population globally. According to the International Treaty on Plant Genetic Resources (Article 9), the farmers and farming communities are the primary custodians of plant genetic possessions and germplasm for food and agriculture (https://www.fao.org/plant-treaty/areas-of-work/farmers-rights/en/).

The use of banana plants and their leaves has been every day in religious ceremonies for ages in India. Traditional foods are prepared by wrapping them in green banana leaves to provide an enriched aroma. Banana leaves are believed to be antibacterial due to their high phenolic compounds. This is traditional knowledge to prepare food in banana leaves by baking or boiling. It needs to be further confirmed by scientific evidence.

The participatory breeding program of roots, tubers, and banana involve farmers and other food processors for variety selection and success of the breeding program through collaborative, multifaceted approaches based on demand in the community and other participants in the food chain. The breeding program of banana is mainly focused on the quality and taste of the final product instead of a high-yield disease-resistance crop. This revolutionary technique is currently being applied in sub-Saharan Africa, and it has profuse potential globally for the success of the breeding program of other major crops.

The application of gene editing tools in agriculture is rapidly expanding worldwide. Since CRISPR/Cas-mediated editing improves crop varieties by modifying their endogenous genome through deletions, insertions, or substitution, or even inserting or replacing a full-length gene at the targeted site precisely, these edited varieties do not contain any external genes from other sources. They do not have to undergo complex and time-consuming biosafety regulations like GMOs for commercialization. The gene-edited crop varieties will be similar to those developed through conventional breeding tools. Gene-edited products of the crops with no external gene integration, especially SDN1 products, are not certified as GMO products in several countries, including Argentina, Chile, Brazil, Colombia, Paraguay, Ecuador, Honduras, Guatemala, Australia, Canada, USA, India, China, Japan, Kenya, and Nigeria (Tripathi et al. 2020).

12 Conclusions

Banana is an excellent fruit full of nutritional value but simultaneously affected by various biotic and abiotic production constraints, especially bacterial, fungal, and viral diseases. The focus of banana researchers is on increasing banana production

by developing disease and pest-resistant varieties by applying transgenic and gene-editing technologies. Few published reports are available on enhancing the nutritional value of the crop, mainly on the golden banana with increased provitamin A. Currently, several bioinformatic tools are available to identify and validate the gene of interest in the banana genome hub, which can be targeted to increase the nutritional value of the crop. Genetic engineering and CRISPR/Cas targeted gene editing are revolutionary technologies that can be adapted to produce banana with the desired nutritional traits.

However, classical genetic engineering, if used, would produce GMOs, which will be regulated and have limited acceptance. In this regard, CRISPR/Cas editing technique becomes an option, especially one which results in producing foreign DNA-free banana. Several countries have formulated and published biosafety guidelines on gene-editing products, especially SDN1 products, which are not regulated as genetically modified products in the absence of a foreign gene or selection marker gene in the final product. As banana is loved and eaten all over, especially by children, plant biotechnologists should explore the possibility of using banana to produce an edible vaccine for future generations.

References

Adedayo BC, Oboh G, Oyeleye SI, Olasehinde TA (2016) Antioxidant and antihyperglycemic properties of three banana cultivars (*Musa* spp.). Scientifica. https://doi.org/10.1155/2016/8391398

Ahmad F, Martawi NM, Poerba YS, de Jong H, Schouten H, Kema GH (2020) Genetic mapping of Fusarium wilt resistance in a wild banana *Musa acuminata* ssp. *malaccensis* accession. Theor Appl Genet 133(12):3409–3418. https://doi.org/10.1007/s00122-020-03677-y

Ainiyah RK, Wahyunindita V, Pratama WN, Pratiwi IA, Utami ESW, Hariyanto S (2020) DNA barcoding: study of banana (*Musa* spp.) wild and cultivars group from East Java inferred by rbcL gene sequences. Ecol Environ Conserv 26:S7–S13. http://www.envirobiotechjournals.com/EEC/26aprilsuppl/EEC-2.pdf

Aman R, Ali Z, Butt H, Mahas A, Aljedaani F, Khan MZ et al (2018) RNA virus interference via CRISPR/Cas13a system in plants. Genome Biol 19:1. https://doi.org/10.1186/s13059-017-1381-1

Arita M, Karsch-Mizrachi I, Cochrane G (2021) The international nucleotide sequence database collaboration. Nucleic Acids Res 49:D121–D124. https://doi.org/10.1093/nar/gkaa967

Bashmil YM, Ali A, Bk A, Dunshea FR, Suleria HA (2021) Screening and characterization of phenolic compounds from Australian grown banana and their antioxidant capacity. Antioxidants 10(10):1521. https://doi.org/10.3390/antiox10101521

Baurens FC, Martin G, Hervouet C, Salmon F, Yohomé D, Ricci S, Rouard M, Habas R, Lemainque A, Yahiaoui N, D'Hont A (2019) Recombination and large structural variations shape interspecific edible banana genomes. Mol Biol Evol 36(1):97–111. https://doi.org/10.1093/molbev/msy199

Biabiany S, Araou E, Cormier F, Martin G, Carreel F, Hervouet C, Salmon F, Efile JC, Lopez-Lauri F, D'Hont A, Lechaudel M (2022) Detection of dynamic QTLs for traits related to organoleptic quality during banana ripening. Sci Hortic 293:110690. https://doi.org/10.1016/j.scienta.2021.11069

Biswas MK, Bagchi M, Biswas D, Harikrishna JA, Liu Y, Li C, Sheng O, Mayer C, Yi G (2020) Genome-wide novel genic microsatellite marker resource development and validation for

genetic diversity and population structure analysis of banana. Genes 11(12):1479. https://doi.org/10.3390/genes11121479

Bolser D, Staines DM, Pritchard E, Kersey P (2016) Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomics data. Methods Mol Biol 1374:115–140. https://doi.org/10.1007/978-1-4939-3167-5_6

Boonruangrod R, Desai D, Fluch S, Berenyi M, Burg K (2008) Identification of cytoplasmic ancestor gene of *Musa acuminata* Colla and *Musa balbisiana* Colla and their hybrids by chloroplast and mitochondrial haplotyping. Theor App Genet 118(1):43–55. https://doi.org/10.1007/s00122-008-0875-3

Borges CV, de Oliveira Amorim VB, Ramlov F, da Silva Ledo CA, Donato M, Maraschin M, Amorim EP (2014) Characterisation of metabolic profile of banana genotypes, aiming at biofortified Musa spp. cultivars. Food chem 145:496–504. https://doi.org/10.1016/j.foodchem.2013.08.041

Brisibe EA, Ubi GM (2020) Microsatellite fingerprinting and analysis of intra-population divergence in morphotaxonomic traits in a large Musa (AAB genome) germplasm. Phys Mole Biol Plants 26(10):1973–1988. https://doi.org/10.1007/s12298-020-00877-0

Brisibe EA, Ubi GM, Ekanem NG (2021) Descriptive and multivariate analyses of morphotaxonomic and yield-related traits in inflorescence dichotomous cultivars of *Musa* species (AAB genome). Genet Resour Crop Evol 68(8):3357–3372. https://doi.org/10.1007/s10722-021-01193-6

Buregyeya H, Tumwesigye KS, Matovu M, Nowakunda K, Namanya P, Tushemereirwe WK, Karamura E, Rubaihayo P (2021) Mode of inheritance for resistance to FOC Race 1 and fruit quality traits in Sukali Ndizi cultivar of Banana. J Plant Breed Crop Sci 13(4):168–176. https://doi.org/10.5897/JPBCS2021.0969

Chen L, Li W, Katin-Grazzini L, Ding J (2018) A method for the production and expedient screening of CRISPR/Cas9-mediated non-transgenic mutant plants. Hortic Res 5:13. https://doi.org/10.1038/s41438-018-0023-4

Chen PJ, Hussmann JA, Yan J, Knipping F, Ravisankar P, Chen P-F et al (2021) Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. Cell 181:5635–5652.e29. https://doi.org/10.1016/j.cell.2021.09.018

Cordenunsi-Lysenko BR, Nascimento JRO, Castro-Alves VC, Purgatto E, Fabi JP, Peroni-Okyta FHG (2019) The starch is (not) just another brick in the wall: the primary metabolism of sugars during banana ripening. Front Plant Sci 10:1–10. https://doi.org/10.3389/fpls.2019.00391

Costa LD, Piazza S, Pompili V, Salvagnin U, Cestaro A, Moffa L et al (2020) Strategies to produce T-DNA free CRISPRed fruit trees via *Agrobacterium tumefaciens* stable gene transfer. Sci Rep 10:20155. https://doi.org/10.1038/s41598-020-77110-1

D'Hont A, Denoeud F, Aury J-M et al (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature 488:213–217. https://doi.org/10.1038/nature11241

Dang TVT, Windelinckx S, Henry IM et al (2014) Assessment of RNAi-induced silencing in banana (*Musa* spp.). BMC Res Notes 7:655. https://doi.org/10.1186/1756-0500-7-655

Davey MW, Gudimella R, Harikrishna JA et al (2013) A draft *Musa balbisiana* genome sequence for molecular genetics in polyploid, inter- and intra-specific Musa hybrids. BMC Genomics 14:683. https://doi.org/10.1186/1471-2164-14-683

Dehery SK, Panda E, Saha PR, Sinha RK, Das AB (2021) Chromosome diversity and karyotype asymmetry analysis in four cultivated triploid and three diploid wild genotypes of Musa from north-East India. Nucleus 64(2):167–179. https://doi.org/10.1007/s13237-020-00334-z

Devarajan R, Jayaraman JK, Somasundaram SM, Ragupathy S, Raman P, Sathiamoorthy K, Subbaraya U (2021) Genetic diversity in fresh fruit pulp mineral profile of 100 Indian *Musa* accessions. Food Chem 361:130080. https://doi.org/10.1016/j.foodchem.2021.130080

Dong C, Wang J, Hu Y, Xiao W, Hu H, Xie J (2022) Analyses of key gene networks controlling carotenoid metabolism in Xiangfen 1 banana. BMC Plant Biol 22(1):1–14. https://doi.org/10.1186/s12870-021-03415-6

Droc G, Larivière D, Guignon V et al (2013) The banana genome hub. Database 2013:bat035. https://doi.org/10.1093/database/bat035

Ekesa B, Nabuuma D, Blomme G, Van den Bergh I (2015) Provitamin A carotenoid content of unripe and ripe banana cultivars for potential adoption in eastern Africa. J Food Compos Anal 43:1–6. https://doi.org/10.1016/j.jfca.2015.04.003

Elayabalan S, Kalaiponmani K, Subramaniam S, Selvarajan R, Panchanathan R, Muthuvelayoutham R et al (2013) Development of *Agrobacterium*-mediated transformation of highly valued hill banana cultivar Virupakshi (AAB) for resistance to BBTV disease. World J Microbiol Biotechnol 9(4):589–596. https://doi.org/10.1007/s11274-012-1214-z

Englberger L, Lyons G, Foley W, Daniells J, Aalbersberg B, Dolodolotawake U et al (2010) Carotenoid and riboflavin content of banana cultivars from Makira, Solomon Islands. J Food Compos Anal 23(6):624–632. https://doi.org/10.1016/j.jfca.2010.03.002

Fungo R, Pillay M (2013) Beta-carotene content of selected banana genotypes from Uganda. Afr J Biotechnol 10(28):5423–5430. https://doi.org/10.4314/ajb.v10i28

Hariyanto S, Ainiyah RK, Utami ESW, Hapsari L (2021) Genetic diversity and network within dessert bananas (*Musa acuminata* Cv. Aa and Aaa) inferred by newly designed matK marker. Int J Conserv Sci 12(2):585–598

Heslop-Harrison JS, Schwarzacher T (2007) Domestication, genomics and the future for banana. Ann Bot 100(5):1073–1084. https://doi.org/10.1093/aob/mcm191

Hu C, Sheng O, Deng G, He W, Dong T, Yang Q et al (2021) CRISPR/Cas9-mediated genome editing of MaACO1 (aminocyclopropane-1-carboxylate oxidase 1) promotes the shelf life of banana fruit. Plant Biotechnol J 19:654–656. https://doi.org/10.1111/pbi.13534

Igwe DO, Ihearahu OC, Osano AA, Acquaah G, Ude GN (2022) Assessment of genetic diversity of *Musa* species accessions with variable genomes using ISSR and SCoT markers. Genet Resour Crop Evol 69:49–70. https://doi.org/10.1007/s10722-021-01202-8

Jeensae R, Kongsiri N, Fluch S, Burg K, Boonruangrod R (2021) Cultivar specific gene pool may play an important role in *Musa acuminata* Colla evolution. Genet Resour Crop Evol 68:1589–1601. https://doi.org/10.1007/s10722-020-01088-y

Jekayinoluwa T, Tripathi JN, Dugdale B, Obiero G, Muge E, Dale J, Tripathi L (2021) Transgenic expression of dsRNA targeting the *Pentalonia nigronervosa* acetylcholinesterase gene in banana and plantain reduces aphid populations. Plants 10:613. https://doi.org/10.3390/plants10040613

Jouneghani RS, Castro AHF, Panda SK, Swennen R, Luyten W (2020) Antimicrobial activity of selected banana cultivars against important human pathogens, including candida biofilm. Foods 9(4):435. https://doi.org/10.3390/foods9040435

Kanazawa K, Sakakibara H (2000) High content of dopamine, a strong antioxidant in Cavendish banana. J Agric Food Chem 48:844–848. https://doi.org/10.1021/jf9909860

Kandasamy S, Aradhya SM (2014) Polyphenolic profile and antioxidant properties of rhizome of commercial banana cultivars grown in India. Food Biosci 8:22–32. https://doi.org/10.1016/j.fbio.2014.10.001

Kanehisa M, Furumichi M, Tanabe M et al (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45:D353–D361. https://doi.org/10.1093/nar/gkw1092

Kang BC, Yun JY, Kim ST, Shin Y, Ryu J, Choi M et al (2018) Precision genome engineering through adenine base editing in plants. Nat Plants 4:427–431. https://doi.org/10.1038/s41477-018-0178-x

Kargar M (2019) Physicochemical and nutritional properties of selected banana (Musa spp.) cultivated in Southeastern US. PhD Thesis. http://hdl.handle.net/10415/6910

Kaur N, Alok A, Shivani S, Kumar P, Kaur N, Awasthi P et al (2020) CRISPR/Cas9 directed editing of lycopene epsilon-cyclase modulates metabolic flux for β-carotene biosynthesis in banana. Fruit Meta Eng 59:76–86. https://doi.org/10.1016/j.ymben.2020.01.008

Khoza M, Kayitesi E, Dlamini BC (2021) Physicochemical characteristics, microstructure and health promoting properties of green banana flour. Foods 10(12):2894. https://doi.org/10.3390/foods10122894

Kleidon J, Brinin A, Paul JY et al (2020) Production of selectable marker gene-free Cavendish banana (*Musa* spp.) using a steroid-inducible recombinase platform. Transgenic Res 29:81–93. https://doi.org/10.1007/s11248-019-00179-6

Kookal SK, Thimmaiah A (2018) Nutritional composition of staple food bananas of three cultivars in India. Am J Plant Sci 9(12):2480–2493. https://doi.org/10.4236/ajps.2018.912179

Kumar GB, Srinivas L, Ganapathi TR (2011) Iron fortification of banana by the expression of soybean ferritin. Biol Trace Elem Res 142(2):232–241. https://doi.org/10.1007/s12011-010-8754-6

Liang Z, Chen K, Li T, Zhang Y, Wang Y, Zhao Q et al (2017) Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. Nat Commun 8:14261. https://doi.org/10.1038/ncomms14261

Liu J, Zhang J, Zhang J, Miao H, Wang J, Gao P et al (2017) Genome-wide analysis of banana MADS-box family closely related to fruit development and ripening. Sci Rep 7:3467. https://doi.org/10.1038/s41598-017-03897-1

Liu Q, Yang F, Zhang J, Liu H, Rahman S, Islam S et al (2021) Application of CRISPR/Cas9 in crop quality improvement. J Mol Sci 22:4206. https://doi.org/10.3390/ijms22084206

Lowder LG, Malzahn A, Qi Y (2018) Plant gene regulation using multiplex CRISPR-dCas9 artificial transcription factors. Methods Mol Biol 1676:197–214

Mertens A, Bawin Y, Abeele SV, Kallow S, Vu DT, Le LT et al (2021) Genetic diversity and structure of *Musa balbisiana* populations in Vietnam and its implications for the conservation of banana crop wild relatives. PLoS One 16(6):e0253255. https://doi.org/10.1371/journal.pone.0253255

Mostafa HS (2021) Banana plant as a source of valuable antimicrobial compounds and its current applications in the food sector. J Food Sci 86(9):3778–3797. https://doi.org/10.1111/1750-3841.15854

Naim F, Dugdale B, Kleidon J, Brinin A, Shand K, Waterhouse P et al (2018) Gene editing the phytoene desaturase alleles of Cavendish banana using CRISPR/Cas9. Transgenic Res 27:451–460. https://doi.org/10.1007/s11248-018-0083-0

Nayar NM (2010) The banana: botany, origin, dispersal. Hortic Rev 36(3):117–164. https://www.cabdirect.org/cabdirect/abstract/20103012849

Ningsih R, Megia R (2019) Folic acid content and fruit characteristics of five Indonesian dessert banana cultivars. Biodiversitas 20(1):144–151. https://doi.org/10.13057/biodiv/d200117

Ntui VO, Tripathi JN, Tripathi L (2020) Robust CRISPR/Cas9 mediated genome editing tool for banana and plantain (*Musa* spp.). Curr Plant Biol 21:100128. https://doi.org/10.1016/j.cpb.2019.100128

Nyine M, Uwimana B, Akech V, Brown A, Ortiz R, Doležel J et al (2019) Association genetics of bunch weight and its component traits in East African highland banana (*Musa* spp. AAA Group). Theor Appl Genet 132:3295–3308. https://doi.org/10.1007/s00122-019-03425-x

Odetayo T, Tesfay S, Ngobese NZ (2022) Nanotechnology-enhanced edible coating application on climacteric fruits. Food Sci Nutr 10:1–19. https://doi.org/10.1002/fsn3.2557

Ortiz R (1997) Occurrence and inheritance of 2n pollen in *Musa*. Ann Bot 79(4):449–453. https://doi.org/10.1006/anbo.1996.0367

Pan C, Wu X, Markel K, Malzahn AA, Kundagrami N, Sretenovic S et al (2021) CRISPR–Act3.0 for highly efficient multiplexed gene activation in plants. Nat Plants 7:942–953

Pandey A, Alok A, Lakhwani D, Singh J, Asif MH, Trivedi P (2016) Genome-wide expression analysis and metabolite profiling elucidate transcriptional regulation of flavonoid biosynthesis and modulation under abiotic stresses in banana. Sci Rep 6:31361. https://doi.org/10.1038/srep31361

Paul J-Y, Khaana H, Kleidon J, Hoang P, Geijskes J, Daniells J (2017) Golden bananas in the field: elevated fruit pro-vitamin A from the expression of a single banana transgene. Plant Biotechnol J 15:520–532. https://doi.org/10.1111/pbi.12650

Pereira A, Maraschin M (2015) Banana (*Musa* spp) from peel to pulp: ethnopharmacology, source of bioactive compounds and its relevance for human health. J Ethnopharmacol 3(160):149–163. https://doi.org/10.1016/j.jep.2014.11.008

Pinar H, Uzun A, Unlu M, Yaman M (2019) Genetic diversity in Turkish banana (*Musa cavendishii*) genotypes with DAMD markers. Fresenius Environ Bull 28(1):459–463

Prakash SP, Chandrasekar V, Subramanian S, Ummar R (2022) The assembled banana dihaploid mitochondrial genome is compact with a high number of gene copies. bioRxiv 2022:1–25. https://doi.org/10.1101/2022.01.13.476214

Roy S, Soni P (2021) Genome editing for biofortification of rice: current implications and future aspects. In: Upadhyay SK (ed) Genome engineering for crop improvement. Wiley, Hoboken, pp 297–313. https://doi.org/10.1002/9781119672425.ch17

Saraswathi MS, Uma S, Ramaraj S, Durai P, Mustaffa MM, Kalaiponmani K et al (2020) Inter retrotransposon based genetic diversity and phylogenetic analysis among the *Musa* germplasm accessions. J Plant Biochem Biotechnol 29:114–124. https://doi.org/10.1007/s13562-019-00519-x

Sardos J, Rouard M, Hueber Y, Cenci A, Hyma KE, van den Houwe I et al (2016) A Genome-Wide Association Study on the seedless phenotype in banana (*Musa* spp.) reveals the potential of a selected panel to detect candidate genes in a vegetatively propagated crop. PLoS One 11(5): e0154448. https://doi.org/10.1371/journal.pone.0154448

Sayers EW, Beck J, Brister JR, Bolton EE, Canese K, Comeau DC et al (2022) Database resources of the National Center for Biotechnology Information in 2023. Nucleic Acids Res 48(D1):D9–D16. https://doi.org/10.1093/nar/gkac1032

Selma S, Bernabé-Orts JM, Vazquez-Vilar M, Diego-Martin B, Ajenjo M, Garcia-Carpintero V et al (2019) Strong gene activation in plants with genome-wide specificity using a new orthogonal CRISPR/Cas9-based programmable transcriptional activator. Plant Biotechnol J 17(9):1703–1705. https://doi.org/10.1111/pbi.13138

Septembre-Malaterre A, Stanislas G, Douraguia E, Gonthier MP (2016) Evaluation of nutritional and antioxidant properties of the tropical fruits banana, litchi, mango, papaya, passion fruit and pineapple cultivated in Réunion French Island. Food Chem 212:225–233. https://doi.org/10.1016/j.foodchem.2016.05.147

Shakirova KM, Ovchinnikova VY, Dashinimaev EB (2020) Cell peprogramming with CRISPR/Cas9 based transcriptional regulation systems. Front Bioeng Biotechnol 8:882. https://doi.org/10.3389/fbioe.2020.00882

Shao X, Wu S, Dou T, Zhu H, Hu C, Hou H et al (2020) Using CRISPR/Cas9 genome editing system to create MaGA20ox2 gene-modified semi-dwarf banana. Plant Biotechnol J 18:17–19. https://doi.org/10.1111/pbi.13216

Someya S, Yoshiki Y, Okubo K (2002) Antioxidant compounds from bananas (*Musa* Cavendish). Food Chem 79:351–354. https://doi.org/10.1016/S0308-8146(02)00186-3

Sunisha C, Sowmya HD, Usharani TR, Umesha M, Gopalkrishna HR, Sxena A (2020) Deployment of stacked antimicrobial genes in banana for stable tolerance against *Fusarium oxysporum* f.sp. cubense through genetic transformation. Mol Biotechnol 62:8–17. https://doi.org/10.1007/s12033-019-00219-w

Telem RS, Wani SH, Singh NB, Nandini R, Sadhukhan R, Bhattacharya S et al (2013) Cisgenics – a sustainable approach for crop improvement. Curr Genomics 14(7):468–476. https://doi.org/10.2174/13892029113146660013

Tripathi L, Ntui VO, Tripathi JN (2020) CRISPR/Cas9-based genome editing of banana for disease resistance. Cur Opin Plant Biol 56:118–126. https://doi.org/10.1016/j.pbi.2020.05.003

Tripathi L, Ntui VO, Tripathi JN (2022) Control of bacterial diseases of banana using CRISPR/Cas-based gene editing. Int J Mol Sci 23:3619. https://doi.org/10.3390/ijms23073619

UniProt Consortium (2021) UniProt: the universal protein knowledgebase. Nucleic Acids Res 49: D480–D489. https://doi.org/10.1093/nar/gkaa1100

Veillet F, Perrot L, Chauvin L, Kermarrec M-P, Guyon-Debast A, Chauvin J-E et al (2019) Transgene-free genome editing in tomato and potato plants using *Agrobacterium*-mediated delivery of a CRISPR/Cas9 cytidine base editor. Int J Mol Sci 20(2):402. https://doi.org/10.3390/ijms20020402

Venkataramana RK, Sampangi-Ramaiah MH, Ajitha R, Khadke GN, Chellam V (2015) Insights into *Musa balbisiana* and *Musa acuminata* species divergence and development of genic

microsatellites by transcriptomics approach. Plant Gene 4:78–82. https://doi.org/10.1016/j.plgene.2015.09.007

Wahyudi DT, Rifliyah K, Uslan (2020) Genome evaluation of banana cultivars based on morphological character and Inter-Simple Sequence Repeat (ISSR) molecular marker. Biodiversitas 21(7). https://doi.org/10.13057/biodiv/d210715

Wang Z, Miao H, Liu J, Xu B, Yao X, Xu C et al (2019) *Musa balbisiana* genome reveals subgenome evolution and functional divergence. Nat Plants 5:810–821. https://doi.org/10.1038/s41477-019-0452-6

Wang J, Wang Z, Jia C, Miao H, Zhang J, Liu J et al (2022) Genome-wide identification and transcript analysis of TCP gene family in banana (*Musa acuminata* L.). Biochem Genet. https://doi.org/10.1007/s10528-021-10100-8

Yadav K, Patel P, Srivastava AK, Ganapathi TR (2017) Overexpression of native ferritin gene MusaFer1 enhances iron content and oxidative stress tolerance in transgenic banana plants. PLoS One 12(11):e0188933. https://doi.org/10.1371/journal.pone.0188933

Zong Y, Song Q, Li C, Jin S, Zhang D, Wang Y et al (2018) Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. Nat Biotechnol 36:950–953. https://doi.org/10.1038/nbt.4261

Apples: Role of Nutraceutical Compounds

Schuyler S. Korban

Contents

S. S. Korban (✉)
Department of Natural Resources and Environmental Sciences, University of Illinois at Urbana–Champaign, Urbana, IL, USA
e-mail: korban@illinois.edu

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_34

Abstract

Apples have long been deemed critical for maintaining a balanced human diet. This is attributed to the apple fruit's contents of fiber, vitamin C, potassium, and various phytochemical secondary metabolites. In fact, apples are known to be excellent sources of various phytochemicals, including secondary metabolites such as polyphenols that have antioxidant properties that protect against oxidative damage. Therefore, these compounds in apple fruit provide additional health benefits to protect against disease in addition to their nutritional value. These antioxidant phytochemical compounds are termed as nutraceuticals. Efforts to investigate, assess, and enhance the nutraceutical content of apple fruits by capitalizing on the collective tools of genetics, genomic selection, genomics, transcriptomics, metabolomics, and other "omics" are contributing to our expanded knowledge of the relatively new field of apple nutraceutomics. However, it is important to keep in mind that there is a complex relationship between nutraceutical content, gene expression, and environmental factors in apple nutraceutomics. This chapter will provide an overview of various nutraceutical compounds present in apple fruits, their roles in benefiting human health and alleviating human health disease, structural and regulatory genes involved in key pathways of secondary metabolites, as well as ongoing efforts and future perspectives in the field of apple nutraceutomics.

Keywords

Antioxidants · Human health · Nutraceutomics · Phenolic compounds · Phytochemical secondary metabolites

1 Introduction

The apple (*Malus* × *domestica* Borkh.), belonging to the family Rosaceae and the tribe Pyreae, is one of the most important temperate fruit crops that is widely grown across the world, and it is widely consumed in multiple forms, including fresh, processed, canned, juiced, and dried. In 2020, the total worldwide production of apples is recorded to be more than 126 million metric tons (Food and Agricultural Organization – FAOSTAT 2022). The top ten apple-producing countries, as of 2020, are China, the USA, Turkey, Poland, India, Italy, Iran, France, Chile, and Russia (FAOSTAT 2022). Although there are thousands of known apple cultivars that are grown worldwide, the commercial apple market is dominated by about dozen cultivars, some that are quite old, such as 'Golden Delicious', 'Delicious', 'Gala', 'Granny Smith', 'Jonathan', 'Braeburn', and 'Fuji', while others that have come along in recent decades such as 'Pink Lady', 'Honeycrisp', 'Envy', and 'Jazz', among others, that have gained commercial importance (O'Rourke 2021; Teh et al. 2021).

The apple, originated in the Tian Shan Mountain region in China, has undergone interspecific hybridization early on, with most of the genome derived from the wild species *M. sieversii* and with some contribution from the crabapple *M. sylvestris* (Khan et al. 2021). The apple is a paleopolyploid but acts as a functional diploid ($2n = 2x = 34$) with a base chromosome number of $x = 17$ that is derived from two successive whole genome duplication (WGD) events of an ancestral genome ($x = 9$) that must have undergone an early WGD event followed by loss of a single chromosome (Khan and Korban 2022; Podwyszyńska and Marasek-Ciołakowska 2021; Celton et al. 2021). This allows for the apple genome to carry most of the genes in duplicate (Li et al. 2019; Podwyszyńska and Marasek-Ciołakowska 2021). The majority of apple cultivars are diploid, but there are some triploid and tetraploid cultivars (Dickson et al. 1992; Podwyszyńska and Marasek-Ciołakowska 2021). The apple is highly heterozygous and self-incompatible and requires cross-pollination to set a good crop (Teh et al. 2021).

It is well known that apple production requires intensive management practices, including training, pruning, cover crop management, irrigation systems, and annual applications of sprays for control of fungal and bacterial diseases, insects, mites, and weeds, thereby contributing to increased costs, particularly in modern high-density orchards (O'Rourke 2021). Thus, it is imperative that decisions are made early in selection of apple cultivars and of rootstocks that are optimal for maintenance in a productive orchard with minimal labor and production costs. Moreover, it is critical that modern apple orchards are tailored for a growing apple industry that is mindful of discerning consumers who are highly aware of the nutritional and health benefits of fresh and processed fruits (Korban 2021).

Apples have long been known to be critical in a balanced human diet as this is due to the fiber, vitamin C, potassium, and antioxidant contents, including various phytochemicals (Brazier 2019). However, these phytochemical compounds in apple fruits that provide extra health benefits such as promoting overall well-being, controlling symptoms, and preventing malignant processes in addition to their basic nutritional values are referred to as nutraceuticals. A medium-sized apple provides the following: 13–20% of a person's daily fiber needs; 9–11% of a person's daily vitamin C needs; and 4% of a person's daily potassium needs (Brazier 2019). Moreover, antioxidant contents contribute to sensory qualities of fresh and processed apple (Mignard et al. 2021) as well as provide benefits to human health (Boeing et al. 2012; Ho et al. 2020; Mignard et al. 2021).

There is increasing evidence that there is a relationship between consumption of fruits and vegetables and reduced risk of various human diseases. In particular, there are many epidemiological studies that have noted that apples are associated with various health benefits. These range from decreased risk of cancer in both women and men, decreased risk of asthma and bronchial hypersensitivity, lower risk for type II diabetes, lower risk of cardiovascular disease (CVD) in women and thrombotic stroke, as well as reduced levels of total cholesterol, particularly LDL ("bad") cholesterol and triglyceride levels, among others (Boyer and Liu 2004; Brazier 2019; Hodgson et al. 2016; Koutsos et al. 2020; Koh et al. 2019). These and other health benefits of apple nutraceuticals will be discussed in more detail in this chapter.

Apples are known to be excellent sources of various phytochemicals, including secondary metabolites such as polyphenols (~110 mg/100 g), as well as fiber (~2–3 g/100 g), and it is these bioactive components that have been investigated for their roles in eliciting potential health benefits (Koutsos et al. 2015, 2020; Koh et al. 2019). Among the group of polyphenols present in apple fruit, flavanols (catechin and procyanidins) account for the largest class (71–90%), followed by hydroxycinnamates (4–18%), flavonols (1–11%), dihydrochalcones (2–6%), and anthocyanins (1–3%), present only in red apples (Koutsos et al. 2015). Based on their chemical structure, phenolic compounds can be subdivided into phenolic acids and flavonoids (Cuthbertson et al. 2012). Within phenolic acids, one of the most relevant in apple is chlorogenic acid, while for flavonoids, these include various compounds including flavonols (quercetin), flavan-3-ols or flavanols (catechin and epicatechin), dihydrochalcone (phloridzin), hydroxycinnamates (coumaric acids, 5′-caffeoyl quinic acid), and anthocyanins (Volz and McGhie 2011). The health protective role of these phenolic compounds is primarily attributed to their redox capacities by allowing for quenching of singlet oxygen molecules and scavenging of free radicals and reactive oxygen species (ROS) (Busatto et al. 2019; Kschonsek et al. 2018). Following ingestion of these polyphenolic compounds, this results in modification of these molecules into other bioactive compounds or this directly allows for interactions with gut microbiota to confer this health benefit (Busatto et al. 2019). Recently, Ichwan et al. (2021) have reported that both flavonoids, in particular quercetin, predominant flavanol in apple peel, and 3,5-dihydroxybenzoic acid, unrelated to flavonoids, but present in apple flesh, are pro-neurogenic. It is reported that quercetin and 3,5-dihydroxybenzoic acid activate precursor cell proliferation, as well as promote cell cycle exit, cellular survival, and neuronal differentiation in brains of test animals, thus promoting adult hippocampal neurogenesis, i.e., brain plasticity (Ichwan et al. 2021). On the other hand, pectin, the main soluble fiber in apples, is reported to influence nutrient transport time, nutrient absorption, and gastric emptying, thus influencing glucose and lipid metabolism (Koutsos et al. 2015; Efimtseva and Chelpanova 2020). Moreover, pectin plays a role in modulating gut microbiota, a key determinant of the bile acid chemical structure, and thus has a signaling potential (Koutsos et al. 2017). Furthermore, fiber appears to help manage blood pressure, which may reduce the risk of cardiovascular disease. Moreover, **vitamin C** is an antioxidant that, alongside other antioxidants, may play a role in protecting some aspects of heart health. Vitamin C may also boost the immune system and help defend the body from infections and diseases. Moreover, potassium helps relax the blood vessels, reducing the risk of high blood pressure and cardiovascular complications. Overall, these critical "health-associated" qualities of apples provide them with advantages that are becoming more readily promoted to expand their marketability.

It is important to point out that levels of polyphenols along with other compounds in apple fruit vary during ripening, as contents of phenolic acids and flavonoids in epicarp and endocarp tissues change during ripening, and these are likely influenced by growing conditions (Stracke et al. 2009; Alberti et al. 2017). However, it is the genetic variability that likely plays a primary role in determining polyphenol levels,

as most of these observed variations are attributed to differences among apple cultivars (Alberti et al. 2017; Kschonsek et al. 2018; McClure et al. 2019).

Assessing and maintaining genetic variability are critical in current concerns over climate change (Boudichevskaia et al. 2020; Mignard et al. 2021). Apple fruit quality and their bioactive compounds are quantitatively controlled traits that are influenced by environmental conditions, such as temperature, solar radiation, and precipitation, as well as management practices, as these contribute to wide variations in accumulation of bioactive compounds (Mignard et al. 2021). In response to increasing concerns over climate change and increasing temperatures, there is an increased interest in developing and selecting for apple cultivars that are resilient to climate change (Boudichevskaia et al. 2020; Mignard et al. 2021).

2 Nutraceutical Content of Apple Fruit

Apple fruits are rich with various groups of secondary metabolites that are known to elicit positive effects on human health (Boyer and Liu 2003–2004). These secondary metabolites consist of a wide set of bioactive phytochemicals including phenylpropanoids, triterpenoids, organic acids, fatty acids, and phenolic compounds (Fig. 1) (Kidoń and Grabowska 2021). Phenylpropanoids contain the flavonoid subclass that includes anthocyanins, pigments responsible for the red color of apple. Triterpenoids are primary components of apple wax (Dashbaldan et al. 2020). The major triterpenoids in apple include oleanolic, betulinic, and ursolic acids along with their derivatives including maslinic, corosolic, euscaphic, pomaceic, and pomolic acids (Sut et al. 2019). Among the most investigated of these phytochemicals are phenolic compounds, as they have been deemed to confer various health benefits (Nezbedova et al. 2021). Some of

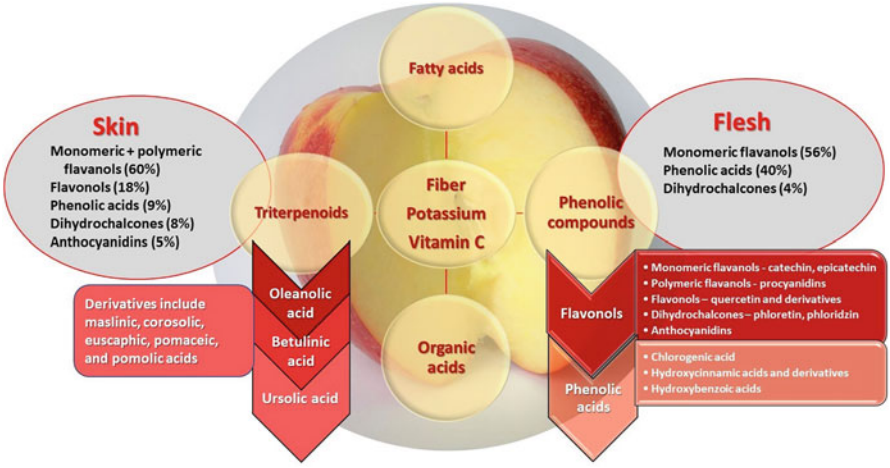


Fig. 1 Groups of phytochemicals present in apple along with average distribution of phenolic compounds in skin and flesh tissues of an apple fruit. (Modified from Nezbedova et al. (2021))

these potential health benefits include anticarcinogenic, anti-inflammatory, and antioxidant properties, among others (Nezbedova et al. 2021; Cosme et al. 2020; Bars-Cortina et al. 2020).

Apples rank second for total content of phenolic compounds. Interestingly, when compared to other fruits, apples have the highest proportion of free phenolics (Boyer and Liu 2004). This renders these free phenolic compounds more available for absorption into the bloodstream. As apples are high in antioxidants, consumption of apples is associated with decreased risk of chronic disease (Boyer and Liu 2004). Apples contain large amounts of various phytochemicals, including phenolic compounds and flavonoids. Levels of these phytochemicals may depend on various factors, including cultivar, harvest, storage conditions, and processing. Moreover, levels of these phytochemicals also vary widely between apple peel and apple flesh. Some of the most well-investigated antioxidant compounds in apples include quercetin-3-galactoside, quercetin-3-glucoside, quercetin-3-rhamnoside, catechin, epicatechin, procyanidin, cyanidin-3-galactoside, coumaric acid, chlorogenic acid, gallic acid, and phloridzin (Boyer and Liu 2004).

Various studies have demonstrated that phenolic compounds in apple contribute up to 22% of phenolic intake in the human diet, and these phenolics in apple fruit are available in conjugated forms as glycosides or esterified carboxylic acids (Nezbedova et al. 2021). However, apple fruits contain more of readily bioavailable free form of phenolic compounds compared to those present in other fruits such as pear, peach, and plum, among others. For example, apple cv. Delicious has the highest level of free forms of phenolics compared to those found in peach, plum, and pear (Imeh and Khokhar 2002).

These antioxidants aid in neutralizing free radicals. Although the terms "reactive oxygen species" (ROS) and "free radicals" are often used interchangeably, they are not synonymous. While ROS is used for those reactive oxygen species whether or not they are free radicals, some ROS are not free radicals, but they are reactive (Stone and Pham 2022). In general, free radicals are reactive molecules that can accumulate due to natural processes and environmental pressures. Thus, when high levels of free radicals accumulate in the human body, they can induce oxidative stress which, in turn, may lead to cell damage. Such damage can contribute to various diseases, including diabetes and cancer. Apples contain a wide range of antioxidants, including quercetin, catechin, phloridzin, and chlorogenic acids (Brazier 2019). Much of the protective qualities of fruits and vegetables have been attributed to various phytochemicals, which are the non-nutrient plant compounds such as carotenoids, flavonoids, isoflavonoids, and phenolic acids. Various phytochemicals are involved in different activities that may help protect against chronic disease. For instance, phytochemicals have been found to inhibit cell proliferation, regulate inflammatory and immune response, and protect against lipid oxidation (Boeing et al. 2012; Ho et al. 2020). In fact, the primary role of phytochemicals is to protect against oxidation. As our living environment is highly oxidative, various processes involved in metabolism may contribute to the production and release of additional oxidants. It is this oxidative damage that plays a role in cardiovascular disease and cancer. Furthermore, with accumulation of such oxidative damage, this will generally

contribute to the aging process. Therefore, it is critical to protect against such oxidative damage to avoid these adverse conditions. It is estimated that there are 10,000 oxidative hits to DNA per cell, per day, in humans (Nezbedova et al. 2021).

The antioxidant compounds commonly found in the apple peel are procyanidins, catechin, epicatechin, chlorogenic acid, ploridizin, and quercetin conjugates. While catechin, procyanidin, epicatechin, and phloridzin are present in the flesh, these are present at lower levels in the flesh than in the peel. Of note, quercetin conjugates are found exclusively in the apple peel, while the only compound that is present at higher levels in the flesh is chlorogenic acid (Nezbedova et al. 2021). In general, it can be concluded that the apple peel contains higher levels of antioxidant compounds. Thus, as the apple peel contains more antioxidant compounds, it can be inferred that the apple peel may have higher antioxidant activity and higher bioactivity than that of the apple flesh, and several studies have confirmed this assumption. Various studies have demonstrated that peeled apples had lower antioxidant activity than apples with peels, as these latter apples are more actively involved in inhibition of cell proliferation, a measure of a compound or fruit extract to inhibit growth of tumor cells. Boyer and Liu (2004) have reported that depending on the cultivar, apple peels contain anywhere from two- to six-fold more phenolic compounds in the peel than in the flesh and two- to three-fold more flavonoids. Furthermore, the antioxidant activity of the peel is also higher than that of the flesh, ranging from two- to six-fold higher, depending on the apple cultivar (Boyer and Liu 2004).

It is known that phenolic compounds in apple fruit belong to one of two major classes, flavonoids and phenolic acids (Fig. 1). Furthermore, flavonoids belong to one of four structural subclasses, namely, flavonols, anthocyanidins, monomeric and polymeric flavan-3-ols (flavanols), and dihydrochalcones (Almeida et al. 2017; Rana and Bhushan 2016). On the other hand, phenolic acids include hydroxycinnamic, hydroxybenzoic, and chlorogenic acids. Overall, among the major phenolic compounds present in apple fruit are monomeric and polymeric flavanols along with chlorogenic acid, while those minor phenolic compounds consist of anthocyanins and dihydrochalcones. It is anthocyanidins that confer the red color of apple fruit, and therefore these compounds are highly abundant in apple cultivars that are red pigmentated, while these levels are either present at low levels or are absent in yellow and green-pigmented apple cultivars (De Paepe et al. 2015; Honda and Moriya 2018; Nezbedova et al. 2021).

As phenolic compound contents in apple are highly variable among different cultivars (Francini and Sebastiani 2013; Brizzolara et al. 2021; Kaeswurm et al. 2022), these contents are also highly variable within tissues of a single fruit, e.g., peel versus flesh (Kim et al. 2019, 2020a; Busatto et al. 2019). For instance, total phenolic contents range between 1157 and 5119 μg/g in peel and 423 and 1534 μg/g in flesh in a group of 12 apple cultivars (Kim et al. 2019). In a recent study, the content of phenolic compounds in seven apple cultivars varied between 112 and 604 mg in 100 g dry weight (or 17–127 mg/100 g fresh weight), while this was significantly higher in peel, ranging between 378 and 1224 mg/100 g dry weight (or 68–285 mg/100 g fresh weight); i.e., the content of phenolic compounds in the dried peel was 2.64- to 4.45-fold higher than that in the dried flesh (Kaeswurm et al. 2022).

These differences in phenolic compound contents between peel and flesh tissues are accompanied by differences in the form of phenolic compounds as well (Brizzolara et al. 2021). For instance, quercetin-3-*O*-galactoside, quercetin-3-*O*-glucoside, catechin, phloridzin, and cyanidin-3-*O*-galactoside are the predominant phenolics in peel (Tsao et al. 2003), while chlorogenic acid, epicatechin, phloridzin, and protocatechuic acids are the predominant phenolics in flesh tissues of apple (Figs. 1 and 2)

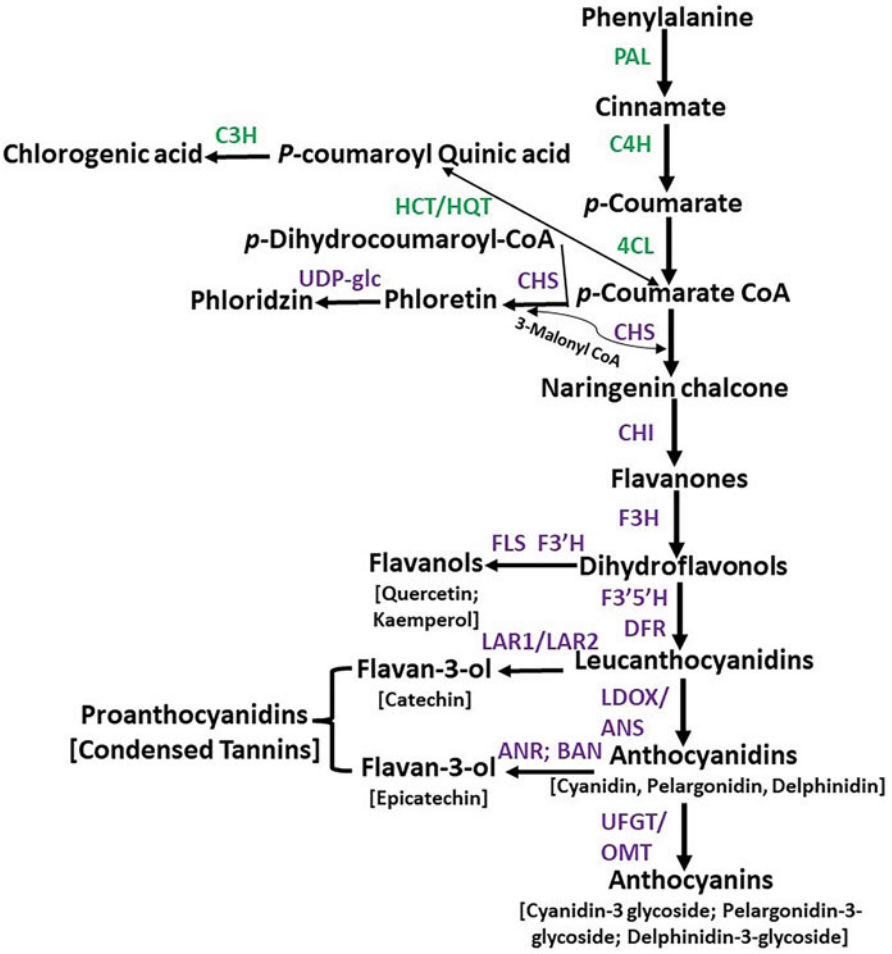


Fig. 2 Schematic diagram of the phenylpropanoid pathway including the flavonoid synthesis and anthocyanin pathways. *PAL* phenylalanine ammonia lyase, *C4H* cinnamate-4-hydroxamate, *4CL* 4-coumarate-coenzyme A ligase, *HQT/HCT* quinate hydroxycinnamoyl/hydroxycinnamoyl CoA shikimate, *C3H p*-coumarate 3-hydroxylase, *UPD-glc* uridine diphosphate glucose, *CHS* chalcone synthase, *CHI* chalcone isomerase, *F3H* flavanone 3-hydroxylase, *F3′H* flavanone 3′-hydroxylase, *FLS* flavonol synthase, *F3′5′H* flavonoid 3′5′-hydroxylase, *DFR* dihydroflavonol-4-reductase, *BAN* BANYULS, *ANR* anthocyanidin reductase, *LAR1/2* leucoanthocyanidin reductase, *LDOX* leucoanthocyanidin dioxygenase, *ANS* anthocyanidin synthase, *UFGT* UDP-glucose flavonoid 3-*O*-glucosyl transferase, *OMT O*-methyl transferase

(Veberic et al. 2005). These observed differences are attributed to genetic diversity, maturity stage, growing conditions, harvest, and storage conditions (Nezbedova et al. 2021; Brizzolara et al. 2021; Alberti et al. 2017; Busatto et al. 2019).

Several studies have confirmed that in general the peel contains about two- to five-fold higher contents of phenolic compounds and of all groups of these compounds, as well as higher contents of total procyanidins and total flavonoids, than those present in the flesh of apple fruit (Łata et al. 2009; Kalinowska et al. 2020; Nezbedova et al. 2021). However, there are some exceptions, e.g., the content of procyanidin B1 in the flesh of 'Gloster', 'Elstar', and 'Gala' is higher compared to that found in the peel (Kalinowska et al. 2014). Similarly, the phloridzin content in the flesh of 'Lodel' is higher than that in the peel (Raudone et al. 2017), as well as the chlorogenic acid content in the flesh of 'Golden Delicious', 'Granny Smith', and 'Idared' is higher in the flesh than that in the peel (Alberti et al. 2017). For some other phenolic compounds, such as quercetin glycosides, these are often found only in the peel (Kschonsek et al. 2018). Furthermore, Kschonsek et al. (2018) have reported that the predominant phenolic compounds found in the flesh are phenolic acids (43%), while flavonols, in particular quercetin and its glycosides, are predominantly present in the peel (72%), and these are not detected in the flesh.

As the apple peel accounts for up to 10% of the weight of a whole fruit, the intake of phenolic compounds following consumption of an apple is not as significant from the peel as that of the flesh (Nezbedova et al. 2021). Earlier, Łata (2007) has reported that the peel of several apple cultivars including 'Braeburn', 'Royal Gala', 'Red Delicious', 'Granny Smith', 'Idared', 'Jonamac', and 'Starking Delicious', among others, accounted for 50% or more of the total content of phenolic compounds in the whole fruit, while the peel of other cultivars, including 'Pilot', 'McIntosh', and 'Prima' contributed less to the total phenolic compounds of the whole apple fruit. Subsequently, Łata et al. (2009) have found that on average, 8%, 24%, 32%, 50%, and 66% of chlorogenic acid, (+)-catechin,(−)-epicatechin, phloridzin, and rutin, respectively, are present in the peel, accounting for 6–8% of the whole apple weight. Moreover, except for chlorogenic acid, on average 50% or more of these phenolic compounds are present in the peel of apple cvs. Granny Smith, Idared, Red Rome, Jonamac, and Gloster, wherein 'Starking Delicious' having the highest (82%) levels of these phenolic compounds in the peel (Łata et al. 2009). Interestingly, anthocyanidins are only present in the peel, while flavonols are primarily found in the peel with only low levels detected in the flesh (Nezbedova et al. 2021). However, it is important to point that some of the above findings seem to vary over years and climate changes (Nezbedova et al. 2021; Mignard et al. 2021).

3 Bioavailability of Nutraceutical Compounds

Any human health-promoting benefits of bioactive compounds of a food component, such as apple fruit, are dependent on absorption, metabolism, and distribution of these compounds within a human body (Marín et al. 2015; Selby-Pham et al. 2017). In fact, bioavailability of phenolic compounds present in apple fruit, once these are absorbed and are available for biological activity, is influenced by pH, enzymatic

activity, solubility, chemical structure, free and bound forms, as well as those synergistic effects with the food matrix (Bondonno et al. 2017; Cosme et al. 2020; Nezbedova et al. 2021; Feng et al. 2021; Di Lorenzo et al. 2021).

Although absorption of phenolic compounds begins in the small intestine (Cosme et al. 2020), it is gut microbia in the large intestine that are responsible for transforming these complex phenolics into metabolites that are easily released and absorbed (Aprikian et al. 2003; Selma et al. 2009; Cosme et al. 2020; Feng et al. 2021). These phenolic compounds of fresh and processed apple can be detected in human plasma and urine (Wruss et al. 2015; Feng et al. 2021; Galvis-Sanchez and Rocha 2016).

Furthermore, as the apple fruit contains various nutrients and phytochemicals, it is also the apple's non-nutrient component, referred to as the food matrix, that plays a critical role in both absorption and bioavailability of these phytochemicals as demonstrated in several studies (Aprikian et al. 2003; Koutsos et al. 2017; Di Lorenzo et al. 2021). Apple fruit has high contents of fiber, pectins, deemed as a complex food matrix (Kaeswurm et al. 2022). It has been reported that uptake of both phenolics-rich apple extract and apple pectin enhances metabolism of gut microbiota in the large intestine, as well as metabolism of lipids more than that of phenolic compounds administered alone, thus demonstrating a positive relationship between phenolic compounds and fiber (Aprikian et al. 2003). Kaeswurm et al. (2022) have found that in a simulated oral digestion study (in vitro and ex vivo), 63–82% of total phenolic compounds are released from the flesh, while 42–58% of phenolic compounds are released from the peel. Furthermore, it has been reported that apple consumption has strong prebiotic effects wherein the fiber content supports bioaccessibility of other beneficial phytochemicals (Efimtseva and Chelpanova 2020; Williams et al. 2017; Starowicz et al. 2020). At the cellular level, fiber is present in the plant cell wall of the apple fruit wherein various phytochemical compounds are reported to bind to plant cell wall components (Williams et al. 2017). On the other hand, consumption of processed apples, in various forms such as juice, sauce, or cooked, impacts the integrity of the plant cell wall and fiber content, thereby modifying bioaccessibility, bioavailability, and phytochemical compound interactions with gut microbiota (Nezbedova et al. 2021).

4 Antioxidant Properties of Apple Nutraceuticals

Recently, there has been an interest in assessing and exploiting those bioactive compounds present in leaf tissues of apple as new raw materials for production of nutraceuticals. Wojdyło et al. (2021) evaluated contents of phenolic compounds, triterpenes, isoprenoids, amino acids, organic acids, minerals, and sugars in leaves and fruits of different cultivars of apple, as well as those of pear and quince. Moreover, they assessed these bioactive compounds for their in vitro enzymatic inhibition of hyperglycemic (α-glucosidase and α-amylase), obesity (pancreatic lipase), cholinesterase (acetylcholinesterase and butyrylcholinesterase), inflammatory (15-LOX, COX-1, and COX-2), and antioxidant capacity (ORAC [oxygen radical absorption capacity], FRAP [ferric reducing antioxidant power], ABTS [2,2-azinobis-(3-ethyl-

benzothiazoline-6-sulfonic acid)]. It is reported that apple leaves have high levels of polyphenolic compounds (about 160.65 g/kg dm), particularly of phloretin-2′-glucose and procyanidins, and these polyphenolics are known to scavenge free radicals and inhibit their production, and they are associated with other biological properties such as anti-inflammatory, cardiovascular, antimicrobial, and anti-cancer, among others. Wojdyło et al. (2021) have found that sugars and organic acids are major components of apple fruits as the sugar/acid ratio, ranging between 8.4 and 35.5, is a critical index of the organoleptic properties of the fruit, as these serve as stimuli for human taste receptors; moreover, organic acids can positively influence microflora in the gastrointestinal tract, thus enhancing their nutritional uptake and health properties. It was found that organic acids in fruits were 1.3–1.4 g/100 g dry weight with dominant acids such as malic acid, followed by quinic and citric acids, while the predominant organic acids in leaves included quinic, malic, malonic, and tartaric acids.

Apple fruit is a rich source of potassium (500–700 mg/100 g dw), serving as one of the dietary minerals that contribute to human health and prevention of human disease (Wojdyło et al. 2021). Furthermore, polyphenolic compounds are present in both apple fruits and leaves; however, apple fruits have high levels of procyanidins, followed by flavonols and then flavan-3-ols, while apple leaves have higher levels of dihydrochalcones, followed by flavonols, and then flavan-3-ols. Additionally, fruits and leaves of different cultivars analyzed exhibited wide differences in flavan-3-ols levels wherein fruits had higher contents than leaves (Wojdyło et al. 2021).

Furthermore, Wojdyło et al. (2021) have reported that the content of dihydrochalcones in different apple cultivars tested ranges between 10.9 and 12.0 mg/100 g dry weight for fruits and from 3155.7 to 3552.7 mg/100 g dry weight for leaves. Recent findings have pointed out that these dihydrochalcones along with other phenolic compounds, such as phloridzin, have a broad spectrum of biological activities including antioxidative, antibacterial, and anti-inflammatory by inhibiting COX-1 and COX-2 activities, anticancer against prostate cancer cells, and immunosuppressive potentials (Wojdyło et al. 2021; Brglez Mojzer et al. 2016). Furthermore, anthocyanin contents in apple fruits are marginal, around 2 mg/100 g dry weight, while anthocyanin contents in leaves are more diverse with significantly higher levels, up to tenfold, than those found in fruits. Anthocyanin compounds strongly absorb ultraviolet light and mainly accumulate in epidermal cells of leaf tissues. The interest in anthocyanin compounds is also attributed not only due to their biological activity, but also to their health-promoting properties (Wojdyło et al. 2021).

Triterpene is present in equal content in leaves and fruits. Of note, leaves of apples and pears most effectively inhibit COX-1, COX-2, α-amylase, and α-glucosidase enzymes (Wojdyło et al. 2021).

5 Health Benefits of Apple Nutraceutical Components

Various studies have demonstrated that phenolic compounds in apple fruit contribute to several biological functions that are beneficial for human health (Nezbedova et al. 2021). Among these various biological functions, it has been reported that these

phenolic compounds contribute to lower incidence of chronic conditions including asthma, pulmonary disease, cardiovascular disease, diabetes, cancer, and obesity, as well as of cognitive decline of normal aging, bone health, and gastrointestinal protection (Hyson 2011; Hosseini et al. 2017; Zhao et al. 2017; Gayer et al. 2019; Hyun and Jang 2016; Koutsos et al. 2015; Mahmud et al. 2018; Zou et al. 2020; Guo et al. 2017; Alam et al. 2018; Martínez-Rodríguez et al. 2020; Ulaszweska et al. 2018).

Furthermore, there are many studies reporting on the chemopreventive and chemoprotective effects of phytochemicals in apple against different forms of cancer. These effects include regulation of cell proliferation, arrest of cell cycle, apoptosis, reactive oxygen species (ROS), and anti-inflammatory activities (Alam et al. 2018; Martínez-Rodríguez et al. 2020; Nezbedova et al. 2021; Nile et al. 2021). Furthermore, these phytochemicals can modulate signaling pathways, thereby exerting epigenetic alterations, such as DNA methylation and expression patterns of microRNAs (miRNAs) (Briguglio et al. 2020).

As mentioned above, intake of apples (and of pears) is reported to reduce incidence of various forms of cancer, such as bladder, breast, lower lung, pancreatic, oral cavity and pharynx, esophagus, larynx, ovary, renal, colorectal, and prostate cancers as reported in various controlled studies conducted in the USA, Europe, China, Mexico, and New Zealand, among other countries (reviewed by Nezbedova et al. 2021). Furthermore, association of apple intake with lower incidence of cancer has also been reported in various observational studies (Fabiani et al. 2016; Reyes-Farias and Carrasco-Pozo 2019; Li et al. 2020). These collective findings are valuable in promoting dietary recommendations; however, these are based on cohort and observational studies and not on clinical intervention trials that are yet to be conducted to confirm direct linkages between apple intake and cancer incidence (Nezbedova et al. 2021).

5.1 The Role of Apple Phenolic Compounds on Cancer

As cancer, in its various forms, is a complex disease that is linked to genetic and molecular controls, cellular and tissue biology, manifestation of disease, and responses to treatment and therapy, it relies on availability of diverse experimental, technological, and computational tools that collectively yield large sets of data related to myriad pathologies of cancer (Nezbedova et al. 2021). Therefore, an integrative concept, referred to as the "hallmarks of cancer," is used to distill the complexity of cancer into a form of science that allows for a more thorough understanding of mechanisms of cancer development and malignant progression. Such hallmarks of cancer correspond to sets of functional capabilities acquired by human cells as they transition from a normal state to an abnormal (neoplastic) growth state; i.e., forming malignant tumors (Hanhan 2022). Studies have reported that phenolic compounds in apple fruit are associated with inhibition of several hallmarks of cancer, such as breast cancer cell proliferation, cell cycle arrest and apoptosis, and growth of cancer cells of lung, prostrate, oral, and colon, among others (Scafuri et al. 2016; Nile et al. 2021; Li et al. 2020; Nezbedova et al. 2021).

In general, it has been reported that phenolic compounds act on carcinogenesis via induction of cell defense systems, including antioxidant enzyme and detoxifying systems. Furthermore, these phenolic compounds are involved in inhibition of anti-inflammatory and anti-cellular growth signaling pathways that lead to arrest of cell cycle and/or to cell death. Thus, these anticancer effects of polyphenolic compounds are capable of altering the epigenome of cancer cells (Blesso 2019; Nasir et al. 2020; Briguglio et al. 2020). It has been demonstrated that dietary natural polyphenolic compounds can elicit anticancer effects via different mechanisms, including cell cycle signaling modulation, removal of anticancer agents, apoptosis, antioxidant enzymes, and cell cycle arrest (Briguglio et al. 2020).

Various studies have reported that apple phytochemicals can inhibit activities of several critical components of cell cycle progression such as growth factors, pyruvate dehydrogenase kinases (PDKs), P21, cyclin-dependent kinases (CDKs), and extracellular protein kinases (ERKs), among others (Nezbedova et al. 2021). Furthermore, phenolic compounds in apple extracts are reported to lower expression of anti-apoptotic genes such as those coding for B-cell lymphoma (Bcl)-2 and p2 proteins while increasing expression of pro-apoptotic genes such as those coding for Bcl-2-associated X (Bax) and p53 proteins (Koh et al. 2019).

Nile et al. (2021) have reported that among the major polyphenolic compounds present in apple pomace, including quercetin-3-glucoside (Q3G), coumaric acid, phloridzin, quercetin, and phloretin, it is Q3G that has the highest antioxidant and anti-inflammatory effects in vitro, as well as elicits significant cytotoxic effects of HeLa cells in a dose- and time-dependent mode. Moreover, flow cytometric analysis has revealed that Q3G induces cell cycle arrest at the S- phase in a time-dependent manner by altering CDK-2, and it induces apoptosis via chromosomal DNA degradation and increased production of reactive oxygen species (ROS) (Nile et al. 2021). In addition, Q3G influences expression of apoptosis-associated proteins by activating caspase-9/-3, upregulating the pro-apoptotic Bax protein, and downregulating anti-apoptosis protein Bcl-2 (Nile et al. 2021).

It is important to point out that these anticancer properties of apples are also attributed to synergistic effects between apple phytochemicals and the food matrix as the food matrix may likely contain other bioactive compounds present in apple extracts (Zhang et al. 2019a; Leng et al. 2018; Nezbedova et al. 2021). Moreover, Han et al. (2019) have reported that extracts of crabapples have more abundant phenolic compounds and higher total polyphenolic compounds, as well as higher antioxidant activities compared to those of apple cv. Fuji. Although all fruit extracts exhibit inhibitory effects on proliferation of different cancer cells, it is crabapple extracts that elicit significantly higher inhibitory effects on colon cancer cells, stomach cancer cells, as well as esophageal cancer cells (Han et al. 2019). Overall, it is reported that red crabapples have higher antioxidant and antiproliferative activities to cancer cells than yellow fruits, as they are richer in phenolic compounds (Han et al. 2019).

It is noteworthy to point out that various bioactive compounds can function as prooxidants, and under conditions of high levels, low pH, and in the presence of metal ions, these compounds can induce ROS production and promote cell death

(Sotler et al. 2019; Kalinowska et al. 2020). In fact, Kalinowska et al. (2020) have reported that the higher content of phenolic compounds in peels of apple fruit correlated with higher pro- and antioxidant activities, as well as in inhibition of lipid peroxidation. In cancer therapy, prooxidants may have a beneficial effect as they serve as cytotoxic agents for fast-growing cells, thus promoting death of cancer cells (Sotler et al. 2019; Fernando et al. 2019). Such prooxidant activities of various phenolic compounds have been associated with induction of apoptosis and arrest of cell cycle arrest of different cancer cells, including skin fibroblast cells (Fernando et al. 2019; Kalinowska et al. 2020). For instance, it has been reported that phloridzin exhibits prooxidant activity, while the antioxidants epicatechin and catechin can act also as prooxidants, and epicatechin induces ROS production in colon cancer cells, thereby activating pro-apoptotic enzymes and inducing apoptosis of these cells (Mendoza-Wilson et al. 2016; Nezbedova et al. 2021). Therefore, phytochemicals in apple can not only have antioxidant effects, but they can also have prooxidant effects for cancer prevention (Nezbedova et al. 2021).

5.1.1 Mechanisms of Anticancer Properties of Apple Phenolic Compounds

Details of the anticancer mechanisms of major phenolic compounds of apple and of other crops investigated in different studies are presented in recent reviews (Nezbedova et al. 2021; Briguglio et al. 2020). Therefore, a summary of the mechanisms of key phenolic compounds involved in the phenylpropanoid biosynthesis pathway is presented herein (Fig. 2).

Among the phenolic compounds that have received much attention, the flavonoid quercetin is reported to promote apoptosis and arrest of cell cycle via its regulation of the proteins p53, GADD45, and AMPK (Guo et al. 2021; Nguyen et al. 2017). Moreover, quercetin induces autophagy in lung cancer and breast cancer cells by reducing activity of the AKT-mTOR pathway (Jia et al. 2018). Furthermore, quercetin inhibits invasion and migration of colorectal and breast cancer cells by inhibiting activity of VEGFR2, thereby inhibiting angiogenesis in prostate, breast, and retinoblastoma cancer cells (Song et al. 2017).

Another group of important flavonoids found in apple include phloretin and phloridzin, a glucoside form phloretin, both metabolized into phloretin glucuronides and phloretin sulfate glucuronides in the human intestine (Choi 2019). Although there is limited knowledge of the bioavailability of phloretin, it is reported that phloretin inhibits proliferation of breast cancer and colorectal cancer cell lines by inhibiting glucose transporter 2 (Glut-2) based on in vitro and in vivo studies (Lin et al. 2016; Wu et al. 2017). Moreover, phloretin also inhibits migration and invasion (Zielinska et al. 2019). Furthermore, it is reported that phloretin inhibits cell proliferation by promoting arrest of the cell cycle, apoptosis, and ROS production and by inhibiting autophagy via mTOR/ULK1 (Min et al. 2015; Kim et al. 2020b; Chen et al. 2020; Xu et al. 2018a, b; Choi 2019). In addition, phloretin supports an anti-inflammatory environment by inhibiting expression of pro-inflammatory molecules including PGE2, IL-8, and AGE (advanced glycation end products) receptors (Zielinska et al. 2019). Recently,

Roy et al. (2022) have reported that a ruthenium-phloretin complex is capable of modulating p53 intervening apoptosis in breast cancer cells. This response is initiated by intrinsic apoptosis, facilitated by Bcl2 and Bax along with down-regulation of the PI3K/Akt/mTOR pathway and coupled with the matrix metalloproteinase-9 (MMP-9) regulated tumor invasive pathways. It is found that such a ruthenium-phloretin chemotherapy could either interrupt, reverse, or suspend succession of breast carcinoma by altering intrinsic apoptosis along with the antiangiogenic pathway (Roy et al. 2022).

Chlorogenic acid in apple fruit is metabolized in the large intestine by gut microbiota, broken down from its aglycone form into various microbial metabolites, including caffeic acid, 3-phenylpropionic acid, hippuric acid, and quinic acid (Koutsos et al. 2017). Although chlorogenic acid has low bioavailability, it is found to inhibit growth of liver and breast cancer tumors in xenograft mice in vivo (Wang et al. 2020; Zeng et al. 2021). Moreover, chlorogenic acid is reported to inhibit cell proliferation, promotes arrest of the cell cycle by modulating the miR-17 family, inhibits invasion and metastasis via downregulation of MMP-9 and MMP-2, and induces apoptosis by binding to annexin and suppressing the NF-κB pathway (Yan et al. 2017; Wang et al. 2020; Zeng et al. 2021; Hou et al. 2017; Sapio et al. 2020).

Among other phenolic compounds in apple are procyanidins, a class of pro-anthocyanidins, consisting of (epi)catechins, particularly B-type procyanidins formed by (+)- and(−)-catechin. These classes of metabolites play critical roles by preventing chronic diseases due to their antioxidant capacity (de la Iglesias et al. 2010); moreover, their bioavailability is determined by their degree of polymerization. For those procyanidins that pass through the stomach unaltered, they are degraded by gut microbiota (especially in the colon), eliciting a prebiotic-like effect (Ou and Gu 2014). More recently, it is reported that catechins and epicatechins, monomeric flavanols are found to be unstable in the gastrointestinal tract, and these are deemed weakly to moderately absorbed based on urinary excretion of human subjects (Almanza-Aguilera et al. 2021). Nevertheless, epicatechins and catechins have been demonstrated to have antioxidant, anti-inflammatory, immunomodulatory, antiviral, and antiallergic effects, and they possess various anticancer properties (Almanza-Aguilera et al. 2021; Nezbedova et al. 2021). In particular, both epicatechins and catechins exhibit inhibitory effects on cancer cell proliferation, induced arrest of the cell cycle via inhibition of CDC25A and upregulation of p21 expression, and lower invasion and migration, as well as induced apoptosis (Pereyra-Vergara et al. 2020; Thomas and Dong 2021; Nezbedova et al. 2021). The anticancer properties of epicatechins and catechin from apples are yet to be fully investigated as most of the previous studies are based on the intake of other food crops.

On the other hand, those procyanidins that are oligomeric flavanols are abundant in apples, and these are known to possess various have anticancer properties. It has been reported that bioavailability of procyanidins is variable, and this is dependent on their structure, as well as the composition of the gut microbiome, wherein low-polymerized procyanidins, such as procyanidin B1, B2, and C1, are more easily absorbed into the small intestine than those highly polymerized procyanidins (Masuda et al. 2018). It is procyanidin B1, B2, and C1 that release metabolites

that may elicit various anticancer effects (Nezbedova et al. 2021). It is reported that procyanidins have higher effects on cancer cell proliferation and apoptosis in vitro compared to other phenolic compounds (Nezbedova et al. 2021). Furthermore, it has been demonstrated, in vivo and in vitro, that procyanidins inhibit cell proliferation via both redox and non-redox regulation of the epidermal growth factor receptor (EGFR) signaling pathway, induce cell cycle arrest by inhibiting the signaling of the epidermal growth factor receptor signaling, induce ROS by inhibiting MAP kinase phosphatase activity and activating ERK1/2 and AMPK, inhibit migration, and promote apoptosis and angiogenesis (Shilpi et al. 2015; Zhu et al. 2020; Daveri et al. 2020; Na et al. 2020; Nezbedova et al. 2021).

Among other bioactive compounds in apple are triterpenoids or pentacyclic triterpenes. These triterpenic compounds are present mainly in apple peel and its associated epicuticular wax. Interestingly, pomolic, euscaphyic, maslinic, and ursolic acids are the most abundant triterpenes in ancient apple cultivars, while in current commercial apple cultivars, it is ursolic acid that is most prevalent, followed by oleanolic acid, and then corosolic acid and betulinic acid (Sut et al. 2019; Butkeviciute et al. 2021). Overall, triterpenic compounds have been reported to have low bioavailability as they are not well absorbed into the intestine (Furtado et al. 2017). These triterpenic compounds are reported to induce apoptosis, inhibit cell proliferation, modify ROS production, and inhibit invasion, metastasis, and angiogenesis via various cellular mechanisms and signaling pathways (Nezbedova et al. 2021). For instance, it has been demonstrated that ursolic acid inhibits breast cancer growth in vitro by inhibiting proliferation, inducing autophagy and apoptosis, as well as suppressing inflammatory responses via the PI3K/AKT and nuclear factor-kappa B (NF-κB) signaling pathways (Luo et al. 2017). On the other hand, oleanolic acid induces osteosarcoma cell apoptosis by inhibiting Notch signaling (Xu et al. 2018). On the other hand, betulinic acid induces apoptosis in human prostate cancer cells via p53 and NF-κB pathways (Shankar et al. 2017).

5.2 The Role of Apple Fruit Nutraceutical Components on Cardiovascular Disease

Consumption of apple has been associated with reduced risk of cardiovascular disease (reviewed by Boyer and Liu 2004). This is primarily attributed to flavonoid intake. In various observational studies, it is reported that women who ingested high amounts of flavonoids experienced 35% reduction in risk of cardiovascular events. Furthermore, consumption of apple resulted in 13–22% reduction in risk of cardiovascular risk in women (Boyer and Liu 2004). In other observational studies, it has been reported that consumption of apples (>71 g/day) by a group of Finnish women resulted in 43% reduction in coronary mortality compared to those who have not consumed apples. On the other hand, the reduction of risk of coronary mortality is 19% in a group of men consuming apples (>54 g/day) compared to those who have not consumed apples (reviewed by Hyson 2011). Furthermore, it has been demonstrated that there is an association of quercetin and apple consumption with

cerebrovascular disease (Boyer and Liu 2004). In other studies, it has been demonstrated that catechin and epicatechin, both readily bioavailable in apples, are highly inversely associated with death from coronary heart disease (Boyer and Liu 2004).

In recent studies, there has been an increased interest in identifying biomarkers associated with cardiovascular risk, particularly focusing on oxidation and lipid metabolism (Hyson 2011). It is known that overexposure to oxidants in the body can contribute to cellular damage, and such oxidative damage may likely disrupt DNA, protein, lipids, and other cellular components via ROS, thus serving to initiate several chronic diseases, such as cardiovascular disease (Hyson 2011). It is reported that dietary antioxidants contribute to the endogenous potential of the body to scavenge ROS and nitrogen-free radicals, thereby directly counteracting lipid peroxidation reactions (Hyson 2011). There are reports demonstrating that apple consumption increases total antioxidant activity in plasma by 64% within a few hours (3 and 6 h) following consumption compared to control (water) treatment (Maffei et al. 2007). Furthermore, it is reported that apple lowers ROS generated by exposure to hydrogen peroxide in lymphocytes isolated from participants in the study. In fact, it is observed that apple significantly protects against DNA damage (highest at 3 h, but this gradually drops by 24 h) in cultured lymphocytes isolated from participants following apple consumption (Maffei et al. 2007). There are several other reports, using animal feeding experiments and in vitro studies, that further demonstrate the role of phenolic compounds in apple and apple products in eliciting antioxidant activities, by inhibiting ROS-induced production of radicals, as well as in modulating lipids and lipid-related processes (Hyson 2011).

5.3 The Influence of Apple Nutraceutical Components on Asthma and Pulmonary Function

Exposure to high and steady levels of oxygen contributes to oxidative damage in lung tissues (Devereux and Seaton 2005). Such oxidant stress activates mediators of pulmonary inflammation that induce asthma (Devereux and Seaton 2005). In particular, oxidative stress can result in various deleterious effects on airway function, such as airway smooth muscle contraction, induction of bronchial hyper-responsiveness (BHR), mucus hypersecretion, epithelial shedding, and vascular exudation (Bowler 2004). Moreover, ROS can activate the transcription factor NF-κB, thus resulting in a cascade of events involving upregulation of transcription of several inflammatory cytokine genes, such as interleukin-6 (IL-6) and leading to influx and degranulation of airway neutrophils (Wood and Gibson 2009). There are a number of studies conducted in different countries that have demonstrated that apple consumption is inversely linked to asthma and it is positively associated with overall pulmonary health (Boyer and Liu 2004; Romieu et al. 2006; Hyson 2011). This reduced incidence of asthma is mostly attributed to apple flavonoids (Romieu et al. 2006). In one study, it is reported that apple and pear intake is positively associated with pulmonary function and negatively associated with chronic obstructive pulmonary disease and that it is catechin that is involved in these observed responses (Hyson 2011).

5.4 The Influence of Apple Nutraceuticals on Anti-inflammatory Responses

Oleanolic acid plays a potential role in treating drug-induced hepatic steatosis via its interaction with liver X receptor alpha and pregnane X receptor, thus reducing ligand-induced lipogenesis (Chen and Lim 2018). Furthermore, it is well known that different triterpene acids from various natural sources have anti-inflammatory effects. For instance, recent studies have reported that maslinic acid has a significant effect on inflammation, and this is partially due to its inactivation of NF-κB (Yap and Lim 2015; Fukumitsu et al. 2016). Furthermore, such anti-inflammatory and anti-arthritic effects of both maslinic and pomolic acids are also attributed to their inactivation of NF-κB (Yap and Lim 2015). Ancient Italian apple cultivars in the *Friuli* Venezia Giulia region are reported to be important sources of such compounds; thus, further investigations should be conducted to assess the potential of these compounds in having anti-inflammatory and glycemic control effects (Sut et al. 2019).

5.5 The Role of Apple Nutraceuticals on Diabetes and Weight Loss

Based on a meta-analysis of various observational studies, it is reported that consumption of apple significantly decreases risk of type 2 diabetes mellitus and body mass index (BMI), in addition to reduced risk of cerebrovascular disease and cardiovascular death (Gayer et al. 2019). Recently, Bondonno et al. (2021) have reported that there is evidence of an inverse association between higher intake of apples and type 2 diabetes for apples, as well as for other fruits including bananas, orange, and other citrus fruits. It is proposed that the beneficial effects of apple consumption (and of other fruits) on glucose regulation and diabetes risk are attributed to multiple factors. These factors include the fruit's low energy intake, low glycemic load, high fiber content, phytochemicals, vitamins, and minerals (Bazzano et al. 2003; Bondonno et al. 2021). Among the phytochemicals in apples, the high content of flavonoids is reported to enhance sensitivity to insulin, thereby decreasing apoptosis, promoting proliferation of pancreatic β cells, as well as reducing muscular inflammation and oxidative stress (Vinayagam and Xu 2015; Kawser Hossain et al. 2016). Furthermore, apple intake may indirectly influence type 2 diabetes risk by either preventing or managing excess adiposity, likely via their higher dietary fiber that contributes to satiety (Guyenet 2019).

5.6 Effects of Apple Nutraceuticals on Various Other Health Diseases

Apple polyphenolic compounds have been reported to inhibit in vitro human low-density lipoprotein (LDL) cholesterol oxidation (Thilakarathna et al. 2013), attenuate Alzheimer's disease (Chan and Shea 2009), and protect against cigarette

smoke-induced acute lung injury (Bao et al. 2013), among other health benefits (Francini and Sebastiani 2013). Koutsos et al. (2020) have reported that apple consumption lowers serum cholesterol and improves cardiometabolic biomarkers in mildly hypercholesterolemic adults. In a recent study, Ichwan et al. (2021) have found that flavonoids, such as quercetin, and 3,5-dihydroxybenzoic acid (not related to flavonoids) are pro-neurogenic, as they activate precursor cell proliferation and promote cell cycle exit, cellular survival, and neuronal differentiation in brains of test animals. Thereby, it is proposed that these compounds are likely involved in promoting adult hippocampal neurogenesis, in particular these compounds are likely involved in brain plasticity. As functional neurons are generated throughout life, these are integrated into existing circuitry; therefore, these compounds are involved in mediating some forms of learning and memory (Ichwan et al. 2021).

However, it is important to keep in mind that bioavailability of apple polyphenolic compounds is a critical factor in their efficacy as disease preventive agents.

5.7 Genetic Diversity of Phytochemical Contents in Apples

Various studies have been undertaken to evaluate the phytochemical contents in different apple cultivars used for the fresh market, as well as for processing – juice, cider, and applesauce, among others (Boyer and Liu 2004; van der Sluis et al. 2001; Escarpa and Gonzalez 1998; Guyot et al. 2002, 2003; Wojdyło and Oszmiański 2020). Studies have reported on the observed variations in total phenolic and total flavonoid contents among different apple cultivars. For example, among ten apple cultivars commonly used for fresh eating, 'Fuji' had the highest total phenolic and total flavonoid contents, while 'Empire' had the lowest (Boyer and Liu 2004). Moreover, among four apple cultivars used for applesauce, namely, 'Rome Beauty', 'Idared', 'Golden Delicious', and 'Cortland', 'Rome Beauty' had the highest phenolic and flavonoid contents, while 'Cortland' had the lowest (Boyer and Liu 2004). On the other hand, 'Idared' had the highest levels of anthocyanins than any of the other cultivars (Boyer and Liu 2004). van der Sluis et al. (2001) reported that among four apple cultivars analyzed, namely, 'Jonagold', 'Golden Delicious', 'Cox's Orange', and 'Elstar', 'Jonagold' had the highest contents of quercetin glycosides, catechins, and chlorogenic acid, followed by 'Golden Delicious'. Escarpa and Gonzalez (1998) reported that among four cultivars analyzed, namely, 'Red Delicious', 'Granny Smith', 'Golden Delicious', and 'Reinata', 'Golden Delicious' had the lowest levels of various flavonoid compounds, while 'Reinata' had the highest level of flavonoids along with 'Granny Smith' and 'Red Delicious'. In addition, Hammerstone et al. (2000) reported that 'Red Delicious' and 'Granny Smith' had the highest content of procyanidins, while 'Golden Delicious' and 'McIntosh' had the lowest.

Stushnoff et al. (2003) assessed the total phenolic content and the antioxidant capacity of a core germplasm collection of the USDA Plant Genetic Resources at Geneva, NY, also maintained at Excelsior (MN), consisting of over 300 *Malus* species, selections, and cultivars. It was found that this collection has a wide

variation for total phenolic content, ranging from 14 to 7181 mg l^{-1}, gallic acid equivalents. In particular, it was observed that the total phenolic content of the commercial cultivars clustered together generally in a narrow and low range, 15–210 mg l^{-1}, the central Asian species *M. sieversii* collection ranged from 100 to 731 mg l^{-1}, and the red-fruited taxa ranged from 151 to 5355 mg l^{-1}. Moreover, 25 taxa representing 10 different *Malus* species and various other genetic backgrounds had the highest total phenolic contents, ranging between 1000 and 7181 mg l^{-1}.

Using a large collection of 80 cultivated apples (*M.* × *domestica*) and 13 accessions of *M. sieversii*, Volz and McGhie (2011) determined levels of flavan-3-ols (catechin + epicatechin); oligomeric procyanidins; flavonols (quercetin 3-rutinoside, quercetin 3-galactoside, quercetin 3-glucoside, quercetin 3-xyloside, quercetin 3-arabinofuranoside, quercetin 3-arabinopyranoside, and quercetin 3-rhamnoside); chlorogenic acid; dihydrochalcones (phloridzin + phloridzin-2-xyloside); anthocyanin [cyanidin 3-O-galactoside (Cy3 gal)]; and total polyphenols in peel and flesh fruit tissues of these 93 apple genotypes for at least 1 year (between 2003 and 2005), grown at a single site in New Zealand. It was reported that genotypic differences for these phenolic compounds accounted for 46–97% of the total variation observed in levels of total polyphenols and for each of the individual phenol groups in both flesh and peel tissues in this germplasm pool. Moreover, it was observed that there were minimal effects for "year" and for "genotype × year" for all phenolic compounds, except for peel flavonols in the cultivated apple, *M.* × *domestica*, and for flesh flavonols in both *M.* × *domestica* and *M. sieversii*, wherein genotypic differences for flavonols accounted for less than 30% of the total variation, which was less than that observed for "genotype × year" interactions. Moreover, levels of total polyphenolic compounds among genotypes ranged between seven- and nine-fold in the flesh and four- and three-fold in peel of *M. sieversii* and *M.* × *domestica*, respectively. In addition, levels of individual polyphenol groups in flesh and peel tissues within each of *M. sieversii* and *M.* × *domestica* ranged between 2- and over a 500-fold. Overall, *M. sieversii*, all accessions originating from Kazakhstan, had higher levels of polyphenolic compounds than those in *M.* × *domestica*. Furthermore, it was found that among all *M.* × *domestica* cultivars and breeding selections originating in New Zealand (since 1990, except for two older cultivars) had lower mean total polyphenols and chlorogenic acid in both flesh and peel tissues than those cultivars originating from other countries, including the USA, UK, France, Germany, Spain, Czech Republic, Netherlands, and Canada.

5.8 Influence of Various Growth and Environmental Factors on Phytochemical Content in Apple

It is important to point out that profiles of phenolic compounds in apple cultivars are highly influenced not only by genetic control but also by growth season, growth period, and geographical location (Wojdyło and Oszmiański 2020). For instance, levels of hydroxycinnamic acids and catechins are high in early developing fruits,

but these levels decrease during fruit growth (Mosel and Herrmann 1974). Similarly, accumulation of flavonoids including flavan-3-ols, dihydrochalcones, and flavonols occurs during early fruiting, and then this drops during fruit growth and maturation (Renard et al. 2007). Likewise, levels of quercetin glycosides, phloridzin, catechins, and chlorogenic acid in 'Jonagold' and 'Elstar' are found to be highest early in the growing season, and these decrease to steady levels during maturation and ripening (Awad et al. 2001a). Furthermore, levels of anthocyanins in both 'Elstar' and 'Jonagold' are high early in the growing season, drop in mid-season, and rapidly increase prior to fruit maturation. Interestingly, such increased levels in anthocyanin content are detected in fruit hanging along the outer periphery of the tree canopy, but not within the inner areas of the tree canopy. Similarly, levels of quercetin glycosides in both 'Jonagold' and 'Elstar' are also found to be higher in fruit hanging along the outer periphery of the tree canopy (Awad et al. 2001b). Thus, sun-exposed fruits of both cultivars have higher levels of both quercetin glycosides and anthocyanins than those of shaded fruits, thereby indicating that exposure to sun light influence increased accumulation of these two phenolic compounds in apple. However, levels of phloridzen, catechin, and chlorogenic acid are not influenced by sunlight. Therefore, it is proposed that light exposure of apple fruit may enhance synthesis and accumulation of particular phytochemicals (Awad et al. 2001b).

Among other factors influencing phytochemical contents in apples is that of cultural management practices such as fertilization. Awad and de Jager (2002a) have reported that application of nitrogen fertilization is associated with drop in levels of anthocyanins, catechins, and total flavonoids in fruit and contributes to lower percentage of blush in peel of fruit of cv. Elstar. On the other hand, calcium fertilization is associated with increased levels in anthocyanins and total flavonoids in fruit of apple cv. Elstar. Furthermore, application of various growth regulators in apple orchards can also influence accumulation of flavonoids and chlorogenic acid as observed on apple cv. Jonagold (Awad and de Jager 2002b). It is reported that application of ethephon highly elevates levels of anthocyanin accumulation, but not of other flavonoid compounds and chlorogenic acid in peel of 'Jonagold', while application of ABG and gibberellic acid (GA3) significantly delayed accumulation of anthocyanin, but not that of other flavonoid compounds and of chlorogenic acid (Awad and de Jager 2002b).

5.9 Effects of Apple Fruit Storage and Processing on Phytochemical Content

It has been reported that phytochemical content in apples is not significantly affected by storage (Boyer and Liu 2004). In an early study, van der Sluis et al. (2001) have observed that levels of quercetin glycosides, phloridzin, and anthocyanin in 'Golden Delicious', 'Jonagold', 'Elstar', 'Red Delicious', and 'Cox's Orange Pippin' do not change following controlled atmosphere (CA) storage for a period of 52 weeks. However, levels of both total catechins and chlorogenic acid slightly decrease in 'Jonagold', while levels of chlorogenic acid remain stable and those of total catechin

slightly decrease in ‘Golden Delicious’. On the other hand, under standard cold storage conditions (0 °C), there is no decrease in levels of chlorogenic acid in all cultivars analyzed, while levels of catechin slightly drop in ‘Elstar’, ‘Cox’s Orange Pippin’, and ‘Golden Delicious’ following 25 weeks of storage (van der Sluis et al. 2001). Interestingly, both forms of storage have no effects on antioxidant activity in any of these tested apple cultivars (van der Sluis et al. 2001).

As apple processing is a major industry, studies have evaluated the influence of processing on phytochemical content. For instance, apple juice from ‘Jonagold’, using straight pressing and pulping, is found to have only 10% of the antioxidant activity of fresh apples. On the other hand, juice resulting from pulp enzyming has only 3% antioxidant activity, and this juice has 58% less catechin, 44% less chlorogenic acid, and 31% less phloridzin, as most of the phenolic compounds are retained in the apple pomace (van der Sluis et al. 2002). It is reported that apple phenolics, especially procyanidins, are found to bind with plant cell wall, which leads to these reduced levels of polyphenol compounds detected in apple juice (Renard et al. 2001). As apple peel contains higher levels of phenolic compounds than that of the flesh, apple peel of ‘Rome Beauty’ is turned into freeze-dried samples, and it is found that these samples have the highest levels of total phenolic and flavonoid contents, even higher than those in fresh peels (Boyer and Liu 2004; Wolfe and Liu 2003). Furthermore, apple peel powder is found to have a strong antioxidant activity and inhibits cell proliferation (Wolfe and Liu 2003). Therefore, apple peel powder may serve as a value-added product in various food products to increase their phytochemical content and antioxidant activity (Boyer and Liu 2004).

In recent studies, it has been found that genotype, tissue type, and cold storage have strong effects on bioactive compound contents in different apple cultivars. It is observed that total phenol content is greatly reduced in flesh (50%) and peel (20%) following cold storage (1 °C for 3 months) in apple cv. Braeburn (clone Hillwell), but not in apple cvs. Golden Delicious (clone B) and Fuji (clone Kiku8) (Carbone et al. 2011). Moreover, phenolic content is found to decrease slightly over storage period (1 °C for 60 days) in commercial apple cvs. Topaz, Pinova, and Pink Lady, as well as in three local (Bosnia and Herzegovina) apple cultivars, namely, ‘Ruzmarinka’, ‘Ljepocvjetka’, and ‘Paradija’ (Begić-Akagić et al. 2011).

5.10 Influence of Applications of Growth Regulator Compounds and Fruit Drying Protocols on Apple Phytochemical Content

Boyer and Liu (2004) evaluated the effects of applying various chemical compounds, used as growth regulators on apple fruit ripening and red color development, as well as the phytochemical content in these fruits. It was observed that ethephon increased anthocyanin production, but it did not increase chlorogenic acid content or the levels of any of other phytochemicals. Moreover, applications of gibberellins and (*S*)-trans-2-amino-4-(2-aminoethoxy)-3-butenoic acid hydrochloride (ABG-3168) decreased anthocyanin production in apple fruit, but these did not influence levels of other phytochemical compounds. Furthermore, application of other chemicals, such as

cycocel, seniphos, shikimic acid, plantacur-E, and galactose, did not have effects on contents of any of the phytochemical contents in apple (Boyer and Liu 2004).

Kidoń and Grabowska (2021) investigated the effects of three fruit drying methods, including convective drying, vacuum-microwave pretreatment with convective drying, and freeze-drying, on the content of phytochemicals, antioxidant activity, color, and sensory attributes of cut cubes of red-fleshed apples, *M. purpurea* or *M. pumila* var. Niedzwetzkyana. Fruits of *M. purpurea* are known to have high contents of anthocyanins both in skin and flesh tissues, rendering the entire apple either red or pink in color, as well as of various phenolic compounds, including chlorogenic acid, catechin, phlorizin, procyanidins, and quercetin derivatives. On average, levels of these phytochemical compounds are threefold higher than those found in standard apple cultivars (Rupasinghe et al. 2010; Wang et al. 2015). Kidoń and Grabowska (2021) observed that following drying, the highest levels of phenolics were detected in freeze-dried apples. In particular, chlorogenic acid was the major phenolic compound, accounting for 60% of all phenolics in both fresh and dried red-fleshed apples, while cyanidin-3-galactoside was the major anthocyanin. However, levels of anthocyanins were markedly lower in these apples following drying.

5.11 Correlations Between Apple Phytochemical Content and Antioxidant Activity

It is important to point out that variations in phytochemical contents among different apple cultivars also influence antioxidant activities. In particular, it has been reported that there is a correlation between levels of phenolic compounds and antioxidant activity (Boyer and Liu 2004). Wojdyło et al. (2008) have determined the composition of phenolic compounds in 67 apple "new" and "old" cultivars along with their antioxidant activities. It is reported that using liquid chromatography-mass spectrometry (LC-MS) analysis for phenolic contents (up to 18 compounds) in this large group of apple cultivars, including those of flavanols (catechin, procyanidin), flavonols, anthocyanins, hydroxycinnamates, and dihydrochalcones, a mean content of total polyphenols ranged between 523 and 2724 mg/100 g dry weight, depending on the apple cultivar. Moreover, it is found that flavanols (catechin and oligomeric procyanidins) account for 80% of polyphenols, followed by hydroxycinnamic acids (1–31%), flavonols (2–10%), dihydrochalcones (0.5–5%), and, in red apples, anthocyanins (1%). Furthermore, it is reported that the highest correlation (r) of total polyphenols and antioxidant activity, based on assays of free-radical scavenging ability, is detected with the 2,2′-azino-bis-3-ethylbenzothiazoline-6-sulphonic acid (ABTS) method ($r = 0.87$), followed by that for the ferric reducing/antioxidant power (FRAP) ($r = 0.84$), and 2,2-diphenyl-1-picrylhydrazyl (DPPH) ($r = 0.80$) methods. Levels of bioactive compounds in this large group of cultivars indicated that these levels are either equal or higher in new (e.g., 'Ozark Gold', 'Julyred', and 'Jester') versus old (e.g., 'Golden Delicious', 'Idared', and 'Jonagold') cultivars, and concentration of procyanidins/flavan-3-ols is the most important contributor to the in vitro antioxidant activity (Wojdyło et al. 2008).

As mentioned above, Stushnoff et al. (2003) evaluated the total phenolic content and the antioxidant activity using the ABTS method of a large collection of *Malus* germplasm. It was reported that the antioxidant activity of a subset of that collection, 103 taxa, correlated well with the total phenolic content ($r = 0.7779$ for juice and $r = 0.7181$ for freeze-dried fruit).

Recently, Mignard et al. (2021) analyzed apple fruit quality and levels of biochemical compounds (peeled fruit, i.e., flesh tissue), including total phenolics content, total flavonoids, vitamin C (ascorbic acid-AsA), and relative antioxidant capacity (using the DPPH method) of 155 accessions, consisting of 99 local (commercial cultivars and traditional landraces) and 56 foreign accessions, grown at the germplasm bank in Spain over a period of 5 years. It was found that total phenolics content widely varied among these accessions and over the years, ranging from 3.3 ('Poma de San Juan', in 2018) to 116.7 ('Camuesa Fina de Aragón', in 2015) mg gallic acid equivalents/100 g fresh weight. On the other hand, the total flavonoid content ranged from 0.7 ('Poma de San Juan', in 2018) to 142.1 ('Camuesa Fina de Aragón', in 2014) mg catechin equivalents/100 g fresh weight. Moreover, levels of vitamin C ranged from 0.4 ('Delgared infel', in 2016) to 13.2 ('Transparente', in 2014) mg AsA/100 g fresh weight. As for the relative antioxidant capacity, this varied from 1.7 ('Poma de San Juan', in 2018) to 44.6 ('Les_1 – MRF 73', in 2014) mg trolox/100 g fresh weight. Over the period of 5 years, mean values of total phenolics content varied from 15.2 ('Biscarri_1 – M 107') to 98.1 ('Camuesa Fina de Aragón') mg gallic acid equivalents/100 g FW, total flavonoid content ranged from 6.0 ('Biscarri_1 – M 107') to 89.0 ('Camuesa Fina de Aragón') mg catechin equivalents/100 g FW, whereas, vitamin C ranged from 1.4 ('Delgared infel') to 5.9 ('Reguard_1 – MRF 53) mg AsA/100 g fresh weight. Finally, mean values of relative antioxidant capacity varied from 5.9 ('Delgared infel') to 30.8 ('Les_1 – MRF 73').

Furthermore, Mignard et al. (2021) have reported that all levels of bioactive compounds and relative antioxidant capacity demonstrated significant differences between local and foreign accessions. Indeed, for all parameters analyzed, foreign cultivars such as 'Deljeni', 'Delorgue Festival', 'Akane', and 'Reineta Gris', among others, had lower values than those of local accessions such as 'Peruco de Caparroso', 'Prau Riu_5', 'Camuesa Fina de Aragón', and 'Les_1 – MRF 73', among others.

Using Pearson's correlation, Mignard et al. (2021) found that relative antioxidant capacity was highly and positively correlated with each of total phenolics content and total flavonoid content, $r = 0.901$ and $r = 0.865$, respectively. Moreover, total phenolics content was significantly and highly positively correlated with total flavonoid content, $r = 0.963$. Interestingly, vitamin C showed a significant and moderate positive correlation with total phenolics content, $r = 0.415$. In addition, significant moderate positive correlations were found between titratable acid and each of total phenolics content and flavonoid content, $r = 0.450$ and $r = 0.464$, respectively.

Using a cluster analysis, Mignard et al. (2021) revealed that foreign cultivars (those not originating from Spain) were concentrated in two groups, while local accessions could not be segregated and had very different profiles. Furthermore, it

was observed that levels of bioactive compounds tended to decrease with higher temperatures, while these levels increased with higher solar radiation; thus, it was concluded that genotypes and climate are major factors influencing variability in metabolite profiles.

Interestingly, Kidoń and Grabowska (2021) used the ABTS assay to assess the antioxidant activity in dried apple fruit cubes of red-fleshed apples (as described in the above section), and they found that apple cubes subjected to convective drying had the highest antioxidant activity, while the vacuum-microwave pretreatment with convective drying samples had the lowest antioxidant activity.

5.12 Genetic Mapping of Phytochemical Content Components in Apple

Efforts to uncover the complex genetic controls of the phenolic compound content of apples have been underway for quite some time, but these efforts have intensified in recent years.

Early genetic mapping efforts of some of metabolites in apple fruits have been undertaken with particular interest in volatile compounds that contribute to aroma. Dunemann et al. (2009) constructed a set of two parental genetic maps for apple cultivars 'Discovery' and 'Prima' using dominant amplified fragment length polymorphism (AFLP) and resistance gene analog (RGA) markers along with a set of 90 multi-allelic simple sequence repeats (SSRs), selected based on their known positions within the *Malus* genome. A total of 18 linkage groups (LGs) was constructed for 'Discovery', while the 'Prima' map consisted of 19 LGs. Hence, 17 homologous LGs corresponding to the basic chromosome number in apple ($n = 17$) were assigned in both parents and aligned with co-dominant SSRs and with segregating AFLPs. A total of 20 volatiles, including various esters, alcohols, terpene α-farnesene, and norisoprenoid β-damascenone, measured using gas chromatography–mass spectroscopy (GC/MS) over a period of 3 years, were analyzed for levels of variability in 'Discovery' and 'Prima'. It was observed that approximately 50 putative quantitative trait loci (QTLs) for a total of 27 different apple fruit volatile compounds were detected via interval mapping by using genotypic data of 150 F_1 seedlings of the mapping population of 'Discovery' × 'Prima' along with phenotypic data obtained by head-space solid phase microextraction gas chromatography. It was reported that QTLs for volatile compounds likely involved in apple aroma were detected on 12 out of 17 apple chromosomes, but these were not evenly distributed. In particular, QTLs were clustered primarily on LGs 2, 3, and 9. A lipoxygenase (LOX) candidate gene, likely involved in volatile metabolism, was mapped on LG 9 and genetically associated with a cluster of QTLs for ester-type volatiles.

With availability of various analytical tools for large-scale analysis of metabolites in an organism, the field of metabolomics has emerged as a powerful tool for pursuing functional genomics studies in various crops, including that of apple (Brizzolara et al. 2021; Han and Korban 2021). Khan et al. (2012a) investigated

the genetic basis of quantitative variations of phenolic compounds in apple fruit. They used a large segregating F_1 population of 'Prima' × 'Fiesta' to map metabolite quantitative trait loci (mQTLs). They conducted untargeted metabolic profiling of peel and flesh tissues of ripe apple fruits using liquid chromatography–mass spectrometry (LC-MS) and detected 418 metabolites in peel and 254 metabolites in flesh tissues (Khan et al. 2012a). Furthermore, using the mapping software MetaNetwork, enabling simultaneous genome-wide screening of numerous traits, they were able to detect and map 669 significant mQTLs, 488 in peel and 181 in flesh tissues. Of particular interest, they identified four linkage groups, namely, LG1, LG8, LG13, and LG16, that contained mQTL hotspots, primarily involved in regulating metabolites belonging to the phenylpropanoid pathway. Subsequently, they used the MapQTL® software to construct an integrated map of both parents, 'Prima' and 'Fiesta', and this map, containing 801 markers and spanning 1348 cM, was used for analysis of annotated metabolites to elucidate the genetics of these annotated metabolites (Khan et al. 2012a). It was observed that a number of quercetin conjugates had mQTLs on either LG1 or LG13. However, an mQTL hotspot with the largest number of metabolites was detected on LG16, wherein mQTLs for 33 peel-related and 17 flesh-related phenolic compounds were located. Furthermore, structural genes of the phenylpropanoid biosynthetic pathway were located by aligning positions of orthologous genes on all 17 chromosomes of apple to DNA sequences of structural genes of *Arabidopsis thaliana* using the genome sequence of apple cv. Golden Delicious (Velasco et al. 2010). Various metabolites that mapped onto an mQTL hotspot on LG16 locus included procyanidins (flavan-3-ols and their polymers), phenolic esters, and flavonol and dihydrochalcone derivatives, among others. All these compounds belong to the phenylpropanoids; thus, it was presumed that this mQTL was controlled by a biosynthetic gene from the phenylpropanoid pathway, or by a transcription factor controlling this pathway. In addition, an apple structural gene coding for leucoanthocyanidin reductase (LAR1), *MdLAR1*, was identified in the mQTL hotspot on LG16, as well as seven transcription factor (TF) genes.

In a subsequent study, Khan et al. (2012b) investigated expression profiles of structural and putative transcription factor genes of the phenylpropanoid and flavonoid pathways during various stages of fruit development in progeny. It was observed that it was only the structural gene *MdLAR1*, located on *LG16*, demonstrated highly significant correlation between transcript abundance and content of metabolites mapped onto the mQTL hotspot. Furthermore, it was observed that seedling progeny of 'Prima' × 'Fiesta', divided into two groups based on "procyanidin dimer II" levels of either high of low content, inheriting either one or two copies of dominant *MdLAR1* alleles (*Mm*, *MM*) had 4.4- and 11.8-fold higher levels of expression, respectively, than progeny inheriting the recessive alleles, *mm* (Khan et al. 2012b). In addition, this observed higher level of expression was associated with a fourfold increase of procyanidin dimer II as a representative metabolite that mapped within the mQTL hotspot. Although expression levels of several other structural genes correlated with expression of various structural genes and with some *bHLH* and *MYB* transcription factor genes, it was only expression of *MdLAR1* that correlated with metabolites mapped within the mQTL hotspot (Khan

et al. 2012b). Therefore, it was proposed that *MdLAR1* was the only candidate gene that could explain the mQTL for procyanidins and flavan-3-ols. However, mQTLs for other phenylpropanoids such as phenolic esters, flavonols, and dihydrochalcones that were detected at the same locus on this map have not been deemed to be dependent on LAR, as their biosynthesis did not involve LAR activity (Khan et al. 2012b). Thus, the dominant allele of *MdLAR1*, promoting increased content of metabolites of potential health benefit, was proposed to be useful in marker-assisted selection in current apple breeding programs, as well as for pursuing cisgenesis (Khan et al. 2012b).

Chagné et al. (2012) analyzed fruits from a population of 'Royal Gala' × 'Braeburn' segregating for phenolic compounds using ultra-high performance liquid chromatography (UHPLC) of extracts derived from fruit peel and flesh. A total of 23 phenolic compounds with varying levels were quantified in apple peel and flesh in 2 separate years. Furthermore, using single nucleotide polymorphic (SNP) markers to genotype individuals in this segregating population, and subsets of these SNPs were used to construct genetic maps for the two parents, 'Royal Gala' and 'Braeburn'. These genetic maps and segregating population were used to detect 79 QTLs for 17 fruit polyphenolic compounds. Of these QTLs, seven QTL clusters were found to be stable across two fruit harvest years, and these included QTLs for content of flavanols, flavonols, hydroxycinnamic acids, and anthocyanins. Following alignment of parental genetic maps with the whole genome sequence of apple cv. Golden Delicious (Velasco et al. 2010), this allowed for screening of candidate genes, coding for enzymes in the polyphenolic biosynthetic pathway, that were co-segregating with these QTLs. As observed by Khan et al. (2012a), the largest cluster of QTLs was located at the top of LG16. Using bioinformatic tools, candidate genes predicted on the basis of their involvement in the polyphenolic biosynthetic pathway were located on the whole genome sequence of cv. Golden Delicious using BLASTN analysis. Furthermore, this co-location was confirmed by genetic mapping of markers derived from gene sequences. It was found that *LAR1* co-located with a QTL cluster for fruit flavanols catechin, epicatechin, and procyanidin dimer, and five unknown procyanidin oligomers identified near the top of linkage group LG 16, while *HYDROXY CINNAMATE/QUINATE TRANSFERASE* (*HCT/HQT*) co-located with a QTL for chlorogenic acid that mapped near the bottom of LG 17. Interestingly, it was hypothesized that the mutation that drove this signal on chromosome 16 was present in the promoter region of *LAR1* and located within a site recognized by transcription factors involved in gene regulation. Furthermore, this mutation did not result in a complete loss of function of *LAR1* (Chagné et al. 2012). It was concluded that *LAR1* and *HCT/HQT* must have likely influenced levels of these phenolic compounds in apple fruit. Therefore, it was proposed that *LAR1* and *HCT/HQT* would serve as useful allele-specific markers for marker-assisted selection of fruit-bearing trees with high content of these phenolic compounds.

Verdu et al. (2014) have capitalized on marker-assisted selection to identify and select new apple cultivars with specific phenolic compounds influencing the taste of cider. Fruit and juice of individuals from a segregating population of a cross between two hybrids X5210 (derived from 'Kermerrien', a well-known French

cider apple cultivar) and X8402 (a desert apple derived from 'Florina' × 'Prima') have been analyzed for various phenolic compounds using a liquid chromatography system. The phenolic compounds including (+)-catechin, (2)-epicatechin, procyanidins B1 and B2, avicularin, hyperin, quercitrin, 5-caffeoylquinic acid, 4-p-coumaroylquinic acid, and phloridzin have been quantified in each 2 years in both fruit and juice. Other compounds (procyanidins B5 and C1, 4-caffeoylquinic acid, isoquercitrin, reynoutrin, ideain, rutin, and phloretin xyloglucoside) have been quantified in some experiments. QTLs have been detected on LG1 for flavanols, LG5 for dihydrochalcones, LG15 for flavonols, and LG16 for mean polymerization degree (DPn) of these phenolic compounds. These QTLs have demonstrated high stability between years and for apple products, fruit, and juice, with high proportion of explained phenotypic variation. Candidate genes under these QTLs for phenolic content were identified *in silico* from the apple genome sequence, and their co-localizations were confirmed by genetic mapping. For example, they identified and mapped four genes homologous to shikimate/quinate O-hydroxycinnamoyl transferase (HCT/HQT) under the QTL confidence interval for the 5-caffeoylquinic acid on the LG17. Likewise, a gene homologous to flavonoid 3′-hydroxylase (F3′H), responsible for the hydroxylation on the third position of the B ring of flavonols, dihydroflavonols, and flavanones, was identified and mapped under the quercetin glycoside cluster. Moreover, a homologue of UDP-glucose 3-glucosyltransferase (UFGT) gene was identified under QTLs for flavanols on LG1. This gene is described to catalyze the formation of anthocyanidins-3-O-β-D-glucoside from anthocyanidins and UDP-D-glucose (Verdu et al. 2014).

Verdu et al. (2014) have confirmed the importance of two regions involved in the biosynthesis of hydroxycinnamic acids on LG14 and LG17. Moreover, other regions of interest include those detected on LG1, LG5, and LG15 for flavanols, dihydrochalcones, and flavonols, respectively. These QTLs are of interest in apple-breeding programs. Furthermore, identification of candidate genes in silico has revealed interesting targets for future studies to better understand the biosynthesis of phenolic compounds. This study not only highlighted QTLs responsible for variability of major phenolic compounds involved in cider organoleptic characteristics but also those for mean polymerization degree of procyanidins. It is reported that these QTLs would aid in understanding the mechanism of procyanidin biosynthesis, which appears to be independent from the synthesis of flavanols.

In another study, a number of major QTLs regulating apple fruit mean L-ascorbate (l-AA), or vitamin C, and total l-AA levels on parental genetic linkage maps of apple cvs. Telamon and Braeburn have been detected (Davey et al. 2006). Moreover, common QTLs were localized to the same region of LGs 6, 10, and 11, accounting for up to 60% of the total observed variation in the segregating seedling population of 'Telamon' × 'Braeburn'. It was proposed that molecular markers for some of these QTL alleles could be used to select for high mean and total contents of l-AA/total l-AA contents. Moreover, a major and highly significant QTL for flesh total l-AA content on LG 17 of the 'Telamon' map was detected that co-localized not only with a QTL for dehydroascorbate (DHA) content but also

with strong QTLs for flesh browning. It is proposed that this QTL on LG 17 may be involved in regulating the redox status of fruit flesh l-AA pool, likely via the activity of polyphenol oxidases (PPOs) or peroxidases (Davey et al. 2006).

Fang et al. (2017) determined the ascorbic acid (AsA) content in mature fruits in a large collection of 30 *Malus* species and 457 accessions, consisting of worldwide cultivars. It was found that AsA concentration ranged from 10.48 to 278.48 μg/g fresh weight, with an average of 46.43 μg/g fresh weight, corresponding to more than 26-fold variation in AsA concentration among all accessions analyzed. Ascorbic acid content ranged from 22.07 to 278.48 μg/g fresh weight in fruits of wild species compared to 10.48 to 131.52 μg/g fresh weight in fruits of apple cultivars. Thereby, this suggested that fruits of wild species had wider variations in ascorbic acid content than fruits of cultivated apples. Furthermore, it was observed that ascorbic acid accumulation in fruit of cultivated apples is rapid during early fruit development, but this markedly decreases during fruit expansion, followed by a slight decline during the mature stage. In an earlier genetic analysis study, Mellidou et al. (2012) reported that four regions on chromosomes 10, 11, 16, and 17 contained stable fruit AsA-QTL clusters, and when AsA metabolic genes were mapped, it was found that within this QTL clusters, these genes co-located with orthologs of GDP-L-galactose phosphorylase (*GGP*), dehydroascorbate reductase (*DHAR*), and nucleobase-ascorbate transporter (*NAT*). Following additional analysis, it was delineated that AsA content in mature apple fruit was mainly controlled by a set of four genes, including two encoding the galactose biosynthetic pathway enzyme genes *MdGGP1* and *MdGGP3*; an ascorbic acid recycling pathway enzyme encoding gene, *MdDHAR3-3*; and a nucleobase ascorbate transporter gene, *MdNAT7-2.* When Fang et al. (2017) evaluated expression of these four genes throughout apple fruit development, it was observed that all four genes were highly expressed in fruits at stages of fruit expansion (highest levels) and maturity stages compared to early fruitlet development at the juvenile stage. Moreover, levels of expression of three genes, *MdGGP1, MdDHAR3-3,* and *MdNAT7-2*, were significantly and negatively correlated with AsA contents in fruits throughout different stages of fruit development. This suggested that low levels of ascorbic acid may induce expression of these three genes during fruit expansion and maturity via a feedback mechanism. However, in young fruitlets, expression of all analyzed genes demonstrated a positive correlation with ascorbic acid content, thereby suggesting that unusual high levels of ascorbic acid detected in early developing fruit was likely due to coordinated contribution of ascorbic acid synthesis and regeneration, as well as of ascorbic acid translocation from leaves (Fang et al. 2017).

A major locus for anthocyanin content in apple flesh has been mapped to linkage group LG 9 (Chagné et al. 2007). Expression studies have determined that this locus is controlled by *MYB10* (Espley et al. 2009). An allelic gene, *MYB1*, also located on LG 9, controls fruit skin color (Zhu et al. 2011). Subsequently, a major QTL controlling red skin coloration on LG 9, found using four segregating New Zealand populations, was screened using an apple 8-K single-nucleotide polymorphism (SNP) array, and a significant SNP marker, ss475879531, was identified (Chagné et al. 2016). This SNP marker was transformed into a marker suitable for use in a real-time PCR assay for the

red skin phenotype (Chagné et al. 2016). It was found that this marker could efficiently predict red skin coloration and it would be useful in marker-assisted selection.

5.13 Genome-Wide Association Studies and Candidate Gene Predictions

McClure et al. (2019) investigated the genetic architecture of polyphenols by combining high performance liquid chromatography (HPLC) data with ~100,000 SNPs using two apple populations, consisting of 136 cultivars evaluated in 2014 and 85 cultivars (from the larger population) evaluated again in 2016, planted in two different randomized blocks. They observed that polyphenol compound contents in fruit, of both peel and flesh, varied widely, up to two orders of magnitude across cultivars, and that much of this variation was both heritable and predicable using genetic markers. Moreover, it was observed that this wide variation was often controlled by a small number of genetic loci with large effects.

Using genome-wide association study (GWAS), McClure et al. (2019) have detected significant genotype-phenotype associations, and the proportion of the phenotypic variance (R^2) explained by the top SNPs ranges from 0.31 to 0.63. Although this relatively high effect size estimates are attributed in part to small sample size, it is suggested that expression of several polyphenolic compounds in apple is under relatively simple genetic control. Furthermore, it is reported that identified markers are in strong linkage disequilibrium (LD) with causal genetic variation that underlies expression of polyphenol compounds. Based on GWAS, it was observed that there was a clear trend for flavan-3-ols and pro-anthocyanidins, wherein there was a large peak on chromosome 16 for catechin, epicatechin, and procyanidin B1, B2, and C1. Interestingly, a highly significant SNP accounted for up to 50% of the observed phenotypic variance. As noted in the above genetic linkage mapping studies of biparental populations, this region on chromosome 16 was identified as a QTL hotspot for catechin, epicatechin, and proanthocyanidins (Khan et al. 2012a, b; Chagné et al. 2012). Although McClure et al. (2019) have detected some slight differences in the location of this most significant SNP across phenotypes, these were within the boundaries of the aforementioned QTL hotspot. Again, as *LAR1* has been identified earlier as a putative candidate gene for this hotspot, it is proposed to catalyze conversion of leucocyanidin to catechin. Moreover, this region is found to also contain several transcription factors of different classes, including *MYB*, *bHLH*, *bZIP*, and *AP2* that could also influence levels of phenolic compounds in apple fruit.

McClure et al. (2019) have found GWAS peaks for various other phenolic compounds including those for flavonol and quercitrin, on chromosome 1. Earlier studies have detected hits for flavonols on chromosome 1 and have proposed that a uridine diphosphate-dependent glycosyltransferase (*UGT*) gene and a flavonoid 3′-hydroxylase (*F3′H*) gene are potential candidate genes underlying this signal (Verdu et al. 2014; Khan et al. 2012a). McClure et al. (2019) have detected a SNP that strongly associated with quercitrin ~44 kb upstream of an apple *UGT* gene

(MD01G1148700). As *UGT*s mediate glycosylation of flavonoids, and quercitrin is produced via glycosylation of the flavonoid quercetin, such glycosylation of secondary metabolites increases both solubility and stabilization of flavonoid compounds; moreover, specific *UGT*s have been identified that glycosylate flavonoids into potent antioxidants, such as phloridzin (Jugdé et al. 2008; Zhou et al. 2017; Kim et al. 2013). Nevertheless, no specific *UGT* has been associated with the formation of quercitrin in apples; thus, it is hypothesized that the GWAS signal detected on chromosome 1 is due to variation in a specific apple *UGT* gene (MD01G1148700) that regulates levels of quercitrin and glycosylation of quercetin (McClure et al. 2019). Further studies will assess whether or not quercetin is in fact associated with this *UGT* gene. Thus, it is anticipated that markers at this locus would be exploited for marker-assisted breeding to enhance the content of quercetin in apple or that genetic variation of antioxidant content will be introduced into new cultivars via gene editing.

McClure et al. (2019) have exploited GWAS for chlorogenic acid, and two significant hits on apple chromosomes 5 and 15 have been detected, thereby suggesting that variation for this trait is controlled by two independent loci. Within this scope of these loci, three promising candidate genes have been identified at these loci including cinnamyl alcohol dehydrogenase (*CAD*; MD05G1089900), caffeoyl-CoA *O*-methyltransferase (*CCOAMT*; MD05G1083900), and 3-dehydroquinate synthase (*DHQS*; MD15G1242600). It is known that both CAD and CCOAMT are enzymes associated with biosynthesis of hydroxycinnamic acids via the phenylpropanoid pathway that also supplies intermediates for synthesis of flavonoids, tannins, and phytoalexins (Hoffmann et al. 2004). As CAD converts cinnamyl alcohol to cinnamaldehyde, it was found that a *CAD* gene was highly expressed in ripening receptacle tissues in strawberry; however, thus far no such *CAD* gene has been characterized in apple (McClure et al. 2019). While CCOAMT is not directly involved in the final step of chlorogenic acid biosynthesis, it is active upstream in its production via conversion of caffeoyl-CoA to feruloyl-CoA, and it has been associated with accumulation of chlorogenic acid accumulation in coffee (Clifford et al. 2017; Hoffmann et al. 2004). As for DHQS, it is reported to be involved in catalyzing key substrates for chlorogenic acid biosynthesis via the shikimate pathway (Maeda and Dudareva 2012). Therefore, it is proposed that all three genes, namely, *CAD*, *CCOAMT*, and *DHQS*, are likely candidates involved in chlorogenic acid production (McClure et al. 2019).

As Khan et al. (2012a) and Chagné et al. (2012) have proposed that *HCT*/*HQT* on chromosome 17 are potential candidate genes for chlorogenic acid, McClure et al. (2019) have not identified SNPs on chromosome 17 that are significantly associated with chlorogenic acid. However, a suggestive GWAS signal is identified on chromosome 17 for chlorogenic acid, and a *HCT*/*HQT* gene (MD17G122510) is found within the suggestive peak on chromosome 17.

Yet another hydroxycinnamic compound, 4-*O*-caffeoylquinic acid, produced significant GWAS hits on chromosomes 3 and 14 (McClure et al. 2019). Although candidate genes for flavonoid 3′-hydroxylase (*F3′H*) or flavonoid 3′,5′-hydroxylase (*F3′5′H*) were proposed, these genes were not located on chromosome 14.

As phenylalanine is converted to *p*-coumaroyl-CoA, with cinnamic acid and *p*-coumaric acid acting as intermediates, it is sequentially catalyzed by phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H), and 4-cinnamoyl-CoA ligase (4CL). McClure et al. (2019) have identified an association signal for 4-*O*-caffeoylquinic acid on chromosome 3; it is proposed that phenylalanine ammonia-lyase (*PAL*; MD03G1121500) may be a candidate gene for this detected signal. PAL, the first enzyme in the phenylpropanoid pathway, is an enzyme that catalyzes the production of cinnamic acid, a precursor to hydroxycinnamic compounds. PAL plays a critical role in controlling the biosynthesis of acyl-quinic acids (Clifford et al. 2017).

In addition, a strong GWAS peak is detected for cyanidin-3-galactoside on chromosome 9, and a strongly associated SNP at this locus is also the most significantly associated SNP with total anthocyanin content in apple (McClure et al. 2019). It is reported that these associations are expected as cyanidin-3-galactoside is the most predominant anthocyanin in apples (Tsao et al. 2003), and several studies have identified a QTL for skin color that is found to co-locate to this genomic region (Zhang et al. 2019b; Amyotte et al. 2017; Chagné et al. 2016). It is reported that a SNP (ss475879531; chr9:33001375) on chromosome 9, useful in predicting skin color in apple breeding programs, is located 666 kb upstream from the most significant SNP identified by McClure et al. (2019). Moreover, Zhang et al. (2019b) have identified a retrotransposon insertion 1 kb upstream of the transcription factor *MYB1* gene (chr9:35,541,127–35,541,721) that likely elicits the red-skinned phenotype. Although this putatively causal allele is located 1.8 Mb downstream from the top GWAS hit, it is found to overlap with a broad GWAS peak detected for both cyanidin-3-galactoside and total anthocyanins.

Therefore, McClure et al. (2019) have identified several SNPs that are deemed as strong candidates for use in marker-assisted breeding. Moreover, it is reported that polyphenol compounds lacking significant GWAS are deemed as predictable using genome-wide SNPs; thus, these may be amenable for breeding using genomic selection (McClure et al. 2019). This study has demonstrated that quite often it is a relatively simple genetic architecture that underlies such observed wide variations in levels of key polyphenolic compounds in apple. These findings offer opportunities for breeding efforts in improving the nutritional value of apples using either marker-assisted breeding or gene editing.

5.14 Transcriptome Expression Profiling of Genes Involved in the Phenylpropanoid Pathway

Busatto et al. (2019) conducted a large transcriptome analysis of 16 genes involved in key steps of the phenolic biosynthetic pathway, beginning with phenylalanine all the way through anthocyanins. These genes included *Phenylalanine ammonia lyase* (*MdPAL*), *Chalcone synthase* (*MdCHS*), *Chalcone-flavone isomerase* (*MdCHI*), *Flavonoid 3-hydroxylase* (*MdF3H*), *Dihydroflavonol 4-reductase* (*MdDFR*), *Anthocyanin synthase* (*MdANS*), and *Anthocyanidin 3-O-glucosyltransferase* (*MdUFGT*).

Moreover, they investigated transcriptomes of genes involved in the synthesis of procyanidins including *leucoanthocyanidin reductase* (*MdLAR*) and *anthocyanidin reductase* (*MdANR*), as well as expression of those major genes involved in the synthesis of flavonols and phoridzin, including *flavonol synthase* (*MdFLS*), *UDP-dependent glycosyltransferase* (*MdUGT*), and *enoyl reductase* (*MdENLR*), respectively. Furthermore, expression of three different *MdUGT* genes, including *MdUGT88F1/4*, *MdMdUGT71K1s,* and *MdUGT71A15*, along with *MdENLR3/5*, that are likely to be involved in the biosynthetic pathway of phloridzin, via glycosylation of phloretin into phloridzin (Zhou et al. 2017), was also investigated. In addition, expression of genes involved in the biochemical pathway of chlorogenic acid and its accumulation and oxidation were also investigated including *p-Coumaroyl ester 3-hydroxylase* (*MdC3H*) and the *Polyphenol oxidase* (*MdPPO*), respectively. This large transcriptome analysis study demonstrated that all these genes were expressed at higher levels in skin tissues than in flesh tissues, collected during early fruit development (74 days after full bloom) and at fruit maturity (harvest time, depending on the genotype), of seven apple genotypes including two wild *Malus* species, *M. baccata* and *M. sieversii*, and five apple cultivars, 'Tyroler Spitzlederer', two clones of 'Golden Delicious' (a smooth skin and a russeted skin clones), 'Cripps Pink', and 'Braeburn'. However, there were some interesting findings wherein, during early fruit development, six genes, namely, *MdCHI*, *MdF3H*, *MdDFR*, *MdANS*, *MdUFGT,* and *MdANR*, had the highest levels of expression in the wild species *M. baccata*, but with no differences in levels of expression between the two tissues (skin and pulp), except for *MdUFGT*, whose expression was much higher in the flesh. On the other hand, transcriptome profiles of *M. sieversii* demonstrated higher levels of gene expression for *MdCHS*, *MdFLS*, *MdLAR,* and *MdENLR3/5*. Moreover, while *MdCHS* was expressed at higher levels in the fresh tissue, *MdFLS* showed higher transcript accumulation in the skin tissue (Busatto et al. 2019).

Transcriptome profiles in cultivated apples demonstrated that *MdPAL*, *MdF3H*, and *McCHI* were highly expressed in skin of the russeted cultivar 'Tyroler Spitzlederer'. Furthermore, the russeted 'Golden Delicious' clone ('Rugiada') demonstrated the highest level of expression of *MdUGT88F1/4* in the skin tissue. These findings supported the role of russeting in the pattern of phenolic compound accumulation in these genotypes due to higher induction of expression of genes in the phenolic biosynthesis pathway of secondary metabolites (Busatto et al. 2019). Furthermore, the majority of genes involved in the phenolic biosynthesis pathway were highly expressed in skin tissues of fruit collected at harvest time in the different genotypes investigated (Busatto et al. 2019). However, at harvest time, transcriptome profiles of all 16 genes were higher in cultivated apples compared to the 2 wild *Malus* species, although *MdCHS* gene in *M. sieversii* was highly expressed. Furthermore, transcriptome profiles of nine genes, namely, *MdPAL*, *MdCHI*, *MdF3H*, *MdANR*, *MdFLS*, *MdENLR3/5*, *MdANS*, *MdC3H,* and *MdUGT88F1/4*, at harvest were significantly higher than during early fruit development, with a 3.8-fold (for *MdUGTF88F1/a*) to 58 (for *MdFLS*). This could be attributed to increased fruit size and skin russeting (Busatto et al. 2019).

5.15 Enhancing Polyphenolic Contents in Red-Fleshed Apples

As noted above, interests in increasing contents of phenolic compounds and their bioavailability have prompted efforts for breeding for red-fleshed apple cultivars. It has been established that the red flesh trait has long been identified in the apple germplasm, particularly in *Malus* species, such as *M. pumila* var. Niedzwetzkyana (van Nocker et al. 2012; Wang et al. 2018). Several studies have reported that a *MYB* TF, *MdMYB10*, has been identified as a key factor responsible for increased accumulation of anthocyanin in several plant tissues and organs, including fruit, via its activation of anthocyanin pathway genes (Espley et al. 2007, 2009, 2013). In particular, owing to the higher anthocyanin contents in red-fleshed apples, compared to white-fleshed ones, particularly with higher levels of cyanidin, along with availability of molecular markers useful for marker-assisted selection for this trait, efforts have been underway to develop red-fleshed apples for its nutraceutical value (Wang et al. 2018).

Busatto et al. (2019) evaluated the phenolic contents of seven white-fleshed apple accessions (described above) along with three red-fleshed apple accessions. As with white-fleshed apples, red-fleshed apple accessions had varying fruit sizes, wherein fruits of cultivated apple (*M.* × *domestica*) and *M. pumila* var. Niedzwetzkyana had an average weight of 80.8 and 75.1 g along with fruit diameter of 6 and 5.8 cm, respectively, while fruit of *M. sylvestris* had an average weight of 17.7 g and fruit diameter of 3.1 cm. All 15 phenolic compound contents, except for anthocyanins, were found in both white- and red-fleshed apples, including neochlorogenic acid, chlorogenic acid, *trans*-piceide, *cis*-piceide, catechin, epicatechin, procyanidin B1, procyanidin B2 + B4, quercetin-3-Rha, kampferol-3-rutinoside, quercetin-3-galactoside + glucoside, isorhamnetin-3-glucoside, rutin, arbutin, and phloridzin. Moreover, levels of chlorogenic acid, epicatechin, and procyanidin B2 + B4 in red-fleshed *M. sylvestris* were significantly higher (90.2, 133, and 16.9 mg/kg fresh weight, respectively), ranging from 1.3- to 4.5-fold, compared to these levels in the two red-fleshed accessions, cultivated apple and *M. pumila* var. Niedzwetzkyana. Comparing the metabolite profiles of white- versus red-fleshed apples, it was observed that the content of phenolic compounds was significantly higher in white-fleshed ones. However, red-fleshed apples had significantly higher anthocyanin contents than white-fleshed ones, although white-fleshed apples had higher contents of all other polyphenols. Furthermore, polyphenolic accumulation in the flesh of white-fleshed domesticated and red-fleshed accessions revealed a fold-change of 17.5, 12.7, and 26.2, respectively, for procyanidins B1 and B2 + B4 and quercetin-3-Rha, and this accumulation pattern is attributed to the fact that the red color is induced by *MdMYB10* TF that activates anthocyanin-related genes in red-fleshed apples (van Nocker et al. 2012; Espley et al. 2007). It is worth pointing out that low levels of phenolic compounds, except for anthocyanins in red-fleshed apples, are noted both in skin and flesh tissues (Busatto et al. 2019). Although it has been widely reported that anthocyanins play important roles as an antioxidants, they have lower levels of bioavailability in comparison to other flavonoids, as they have been found to either degrade rather quickly or are rapidly absorbed and excreted in humans (McGhie and Walton 2007; Fernandes et al. 2014).

5.16 Structural and Regulatory Genes Controlling Critical Nutraceutical Biosynthesis Pathways

5.16.1 The Phenylpropanoid Pathway

The phenylpropanoid biosynthesis involves several structural genes, including those coding for phenylalanine ammonia-lyase (*PAL*), cinnamate-4-hydroxamate (*C4H*), and 4-coumarate–CoA ligase (*4CL*) (Fig. 2; Liu et al. 2021). This is followed by early biosynthesis genes (EBGs) of the flavonoid biosynthesis, including chalcone synthase (*CHS*), chalcone isomerase (*CHI*), flavanone 3-hydroxylase (*F3H*), flavanone 3-′-hydroxylase (*F3′H*), and flavonoid 3′5′-hydroxylase (*F3′5′H*) (Henry-Kirk et al. 2012; Davies et al. 2020). This is followed by late biosynthesis genes (LBGs) of the flavonoid biosynthesis pathway including dihydroflavonol 4-reductase (*DFR*), anthocyanidin synthase (*ANS*)/leucoanthocyanidin dioxygenase (*LDOX*), UDP-glucose flavonoid 3-*O*-glucosyl transferase (*UFGT*), flavonol synthase (*FLS*), leucoanthocyanidin reductase (*LAR1*/*LAR2*), anthocyanidin reductase (*ANR*), glycosyltransferases (*GT1/GT2*), quinate hydroxycinnamoyl (*HQT*)/hydroxycinnamoyl CoA shikimate (*HCT*), and *p*-coumarate 3-hydroxylase (*C3H*), along with other modifying genes, such as methyltransferase (*MT*), O-methyltransferase (*OMT*), and anthocyanin transferase (*AT*) (Han et al. 2007, 2012; Henry-Kirk et al. 2012; Davies et al. 2020; Liu et al. 2021).

In addition, there are four structural genes, namely, transparent testa 10 (*tt10*), transparent testa 12 (*tt12*), transparent testa 13 (*tt13*), and transparent testa 19 (*tt19*), encoding polyphenol oxidase (PPO), a secondary transport factor of the MATE (multidrug and toxic compound extrusion) family, H (+)-ATPase, and glutathione S-transferase (GST), respectively. These proteins play significant roles in modification, transport, and oxidation of anthocyanidins (Liu et al. 2021).

5.16.2 The Flavonoid Biosynthesis Pathway

The flavonoid biosynthesis pathway has been extensively investigated in several plant species (Brueggemann et al. 2010; Liu et al. 2021). It is reported that regulation of this pathway occurs mostly at the transcriptional level of those structural genes encoding enzymes. It has been found that two major groups of transcription factors (TFs) are found to be involved in all investigated plant species, and these include the basic helix-loop-helix (bHLH) and R2R3 myeloblastosis (MYB) family proteins. BHLH proteins can typically regulate target genes in multiple branches of the flavonoid biosynthesis pathway (Brueggemann et al. 2010), while specificity for regulation of a single branch is conferred by MYB factors (Brueggemann et al. 2010). Although transcriptional control of the flavonoid pathway occurs mainly with these two major groups of TFs, additional regulators are required for proper expression of structural genes (Fig. 3).

As reported above by Khan et al. (2012b) and Chagné et al. (2012), an *LAR* gene within a major QTL is deemed as a strong candidate controlling the accumulation of both flavanols and procyanidins. However, Brueggemann et al. (2010) have observed that an apple WD40-repeat gene (*MdTTG1*), a homolog of *Arabidopsis* TRANSPARENT TESTA GLABRA1 (TTG1), activated a promoter of the

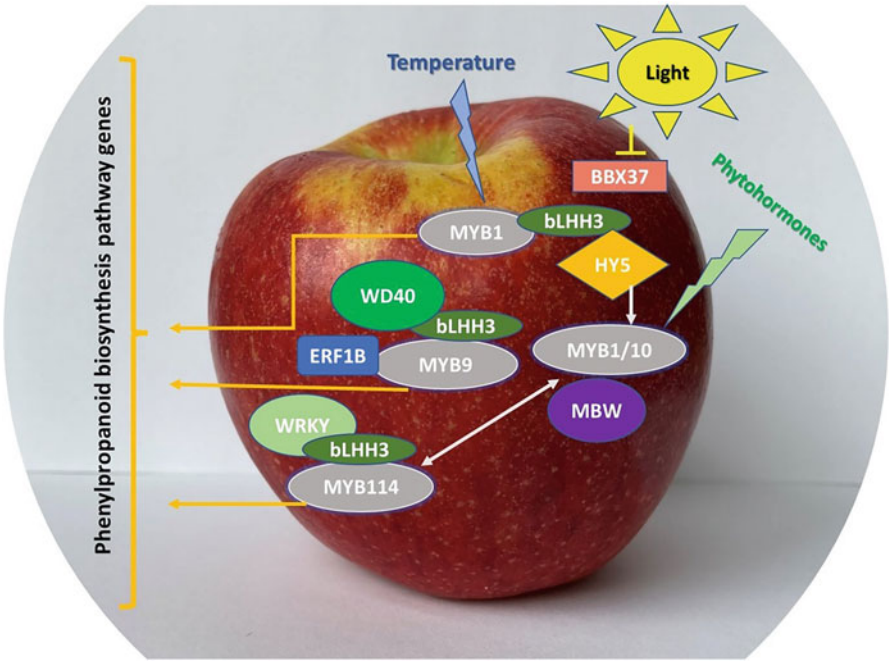


Fig. 3 A simplified schematic diagram of some of the key regulatory factors involved in the phenylpropanoid biosynthesis pathway. *BBX* B-box, *bHLH* basic-helix-loop-helix, *ERF* ethylene response factor, *HY5* ELONGATED HYPOCOTYL 5, *MYB* v-*myb* avian myeloblastosis viral oncogene homolog, *WD40* WD repeat protein, *WRKY* WRKYGQK domain, *MBW* MYB-bHLH-WD40

Arabidopsis *BANYULS* gene, *AtBAN,* along with *Arabidopsis* TT2 and TT8 in *A. thaliana* protoplasts. This suggests that proanthocyanidin (PA) accumulation in apple is likely regulated at the transcriptional level although no TFs in apple have been yet identified to be involved in regulation of PA biosynthesis.

Earlier, Han et al. (2012) investigated the functionality of the anthocyanidin reductase (*ANR*) gene family in apple, consisting of a single *MdANR1* gene on chromosome 10 and two allelic *MdANR2* genes, *MdANR2a* and *MdANR2b*, on chromosome 5. Liao et al. (2015) reported on the role of *LAR* genes in proanthocyanidins (PA), also referred to as condensed tannins, biosynthesis in apple. Expression profiles of both *LAR* and *ANR* genes were investigated in fruit skin of a single cultivated apple and three crabapples. Transcript levels of *LAR1* and *ANR2* genes were significantly correlated with contents of catechin and epicatechin, respectively, thus suggesting that these genes played active roles in PA synthesis. However, unexpectedly, transcript levels for both *LAR1* and *LAR2* genes were almost undetectable in two crabapples accumulating both flavan-3-ols and PAs. This finding contradicted an earlier finding that *LAR1* gene was a strong candidate involved in regulation of metabolite accumulation of epicatechin and PAs in apple. Moreover, ectopic expression of the apple *MdLAR1* gene in tobacco was found to

suppress expression of the late genes in the anthocyanin biosynthesis pathway, thereby leading to resulting loss of anthocyanin in flowers. Furthermore, a decline in PA biosynthesis was also observed in flowers of transgenic tobacco plants overexpressing the *MdLAR1* gene, which was likely attributed to reduced levels of expression of both tobacco genes *NtANR1* and *NtANR2* genes. This study confirmed the in vivo function of the apple *LAR1* gene (Liao et al. 2015).

5.16.3 The Anthocyanin Biosynthesis Pathway

As for anthocyanin biosynthesis, the role of regulatory genes is of critical importance. Many genes involved in regulation of the anthocyanin biosynthetic pathway consist of groups of TFs, including those of the MYB family, the bHLH family, and the tryptophan-aspartic acid repeat (WDR) family (Fig. 3) (Tian et al. 2017; Liu et al. 2021). Often, members of these three families of regulatory genes are dependent on the MYB-bHLH-WD40 (MBW) complex to elicit their roles (Hichri et al. 2010; Liu et al. 2021; Li et al. 2022). Those regulatory genes involved in upregulation of the anthocyanin pathway are members of a subclade that includes the *PRODUCTION OF ANTHOCYANIN PIGMENT1* (*PAP1*) gene, and when overexpressed in *Arabidopsis*, this results in accumulation of anthocyanins (Borevitz et al. 2000). Anthocyanin-regulating MYBs have been isolated from various plant species, including apple, among other fruit crops in the Rosaceae family (Lin-Wang et al. 2010; Vimolmangkang et al. 2013; Chagné et al. 2013; Han et al. 2010). It is observed that proanthocyanidins are synthesized from epicatechin, which in turn is catalyzed by anthocyanidin reductase (ANR), transported, and polymerized (Tian et al. 2017). Anthocyanins are synthesized on the cytoplasmic surface of the endoplasmic reticulum, and stable anthocyanins are formed via different modifications, including glycosylation, methylation, and acylation, and then transported into the vacuole where they accumulate (Tian et al. 2017; Liu et al. 2021).

Plunkett et al. (2019) have reported that accumulation of anthocyanins in plants seems to be linked to biotic and abiotic stress, with levels of anthocyanin increasing in response to these stresses. Studies on regulation of anthocyanin production have identified both biosynthetic pathways genes and major regulating TFs, comprising the MYB-bHLH-WD40 (MBW) complex (Baudry et al. 2004; Henry-Kirk et al. 2012; Albert et al. 2014). This complex binds promoter regions of anthocyanin biosynthetic genes and *MYB10* to enhance transcription, as demonstrated in various plant species including apple (Espley et al. 2009; Chagné et al. 2013; Plunkett et al. 2019). Those MYBs regulating accumulation of anthocyanin in apple have been well investigated (Allan et al. 2008; Ban et al. 2007; Espley et al. 2009; Takos et al. 2006). It is found that the alleles *MYBA* and *MYB1* are identical, sharing 98% sequence homology with *MYB10*, differing by three amino acids in the open reading frame (Ban et al. 2007). Furthermore, the R2R3 binding region of these genes and anthocyanin-related MYBs in other plant species is found to be highly conserved (Ban et al. 2007). It has been observed that MYB10, bHLH3/33, and a WD40 protein TTG1 in apple form an MBW complex that regulates anthocyanin levels in other plant systems (Allan et al. 2008; An et al. 2012).

It is also important to point out that repressor genes also play an important role in the regulation of anthocyanin production, and these repressors have been identified in various plant species, such as *Arabidopsis*, petunia, strawberry, and apple, among others (Plunkett et al. 2019). For instance, two repressors have been reported in petunia: MYB27 (R2R3-MYB), a R2R3-type MYB that interferes with the MBW complex by preventing formation or incorporation and converting the complex into a repression complex, and MYBx (R3-MYB), an R3-type MYB that competes for interaction with the bHLH component of the MBW complex (Albert et al. 2014). MYB27 contains an ethylene-responsive element binding factor [ERF]-associated amphiphilic repression domain (EAR) and a repression TLLLFR motif, both involved in conferring the capability of anthocyanin repression (Albert et al. 2014). Lin-Wang et al. (2011) have identified a family of MYB repressors of anthocyanin in apple, belonging to the R2R3 MYB family, following heat treatment, and these repressors are characterized by an EAR motif. As Lin-Wang et al. (2011) have demonstrated that heat reduces *MYB10* expression in apple skin, Xie et al. (2012) have demonstrated that low temperature induces binding of bHLH3 to *MdMYB1* and to upregulating promoters of *MdDFR* and *MdUFGT* genes of the anthocyanin biosynthesis pathway (Fig. 3).

Thus, it is clear that genes controlling the phenylpropanoid biosynthetic, flavonoid, and anthocyanin pathways in plants are mainly regulated by transcriptional changes (Plunkett et al. 2019; Liu et al. 2021).

As mentioned above, anthocyanins are important secondary metabolites, belonging to flavonoids (polyphenols) (Liu et al. 2021). Thus far, over 20 anthocyanins have been identified in nature, and these are derived from the six most common anthocyanins, namely, cyanidin (Cy), peonidin (Pn), pelargonidin (Pg), malvidin (Mv), delphinidin (Dp), and petunidin (Pt). Anthocyanins are present along with various monosaccharides, including glucose, rhamnose, galactose, and xylose, as well as with disaccharides consisting of rhamnose, gentian disaccharide, and sophora disaccharide to form glycosides. Furthermore, anthocyanins consist of α-phenylbenzopyran cations, and these are primarily composed of C6(A)-C3 (C)-C6(B) carbon skeleton structures. Depending on presence of hydroxyl groups at the 3′ and 5′ of the B-ring carbon structure, along with methoxylation, this is used to distinguish among the six anthocyanins. Thus, methylation and hydroxylation modifications at different positions of the B-ring of the molecule contribute to the development of different colors of anthocyanins.

In addition to the roles of heat and cold temperatures in regulating anthocyanin accumulation, light conditions are also required for anthocyanin accumulation in many apple cultivars, particularly following exposure to ultraviolet-B (UV-B) irradiation (Ban et al. 2007; Jakopic et al. 2009; Vimolmangkang et al. 2014; Bai et al. 2014).

R2R3-MYB TFs play critical roles in the regulation of the anthocyanin biosynthesis pathway as these can directly regulate expression of related genes, thereby contributing to tissue-specific anthocyanin accumulation (Liu et al. 2021). It has been well documented that BHLH TFs are necessary for activities of R2R3-MYBs, primarily for stabilizing the MYB complex or for promoting its transcription (Liu et al. 2021). As an

example, some MYB TFs, such as MdMYB1, MdMYB9, MdMYB10, and MdMYB114, can promote apple fruit color development via interactions with bHLH3 and WD40 (Ban et al. 2007; Jiang et al. 2021). MdMYB1 is initially found in apple fruit skin, and its product participates in photoinduction by activating the transcription activity of promoters of the structural genes *MdDFR* and *MdUFGT* (Xie et al. 2012). It has been well documented that transcript levels of *MdMYB1* are positively correlated with anthocyanin accumulation and expression of structural genes as promoters of *MdDFR* and *MdUFGT* have light-response elements, such as ACGTs and MRE or MRE-like sequences (Takos et al. 2006; Hartmann et al. 2005). Zhang et al. (2022) have reported that mdm-mir858, an miRNA with multiple functions in plant development, negatively regulates proanthocyanidin accumulation by targeting *MdMYB9/11/12* in the peel of apple fruit.

Members of the WD40 protein family have 4–10 random WD repeat domains, consisting of 40 amino acid sequences ending with tryptophan (W) and aspartic acid (D). Among the first WD40 proteins isolated from apple is MdTTG1, and this can interact with MdbHLH3 and MdMYB9 to control expression of downstream structural genes (Brueggemann et al. 2010). Some TFs such as apple MYB16 can negatively regulate anthocyanin biosynthesis by hindering formation of the MBW complex (Xu et al. 2017).

Although many TFs have been found to be involved in the regulation of anthocyanin biosynthesis, the role of light in regulating anthocyanin biosynthesis is still under study. Various investigations of specific members of the *BBX* (B-BOX) protein gene family, belonging to the zinc-finger transcription factors, in apple have revealed that UV-B light in particular upregulates *MdCOL11/BBX33* expression in a temperature-dependent manner (Liu et al. 2021). *MdBBX33* is a close homolog of *AtBBX22/LZF1 (LIGHT-REGULATED ZING FINGER PROTEIN 1)*, also known as *STH3* (Salt Tolerance Homolog 3) and *DBB3* (Double B-Box zinc finger 3) (Bai et al. 2014). Furthermore, overexpression of apple *BBX33* in transgenic *Arabidopsis* lines resulted in increased anthocyanin accumulation and revealed that expression profiles of both *BBX33* and *MYB10* are related during different temperature and light regimes (Bai et al. 2014). It is demonstrated that BBX33 upregulates the *MYB10* promoter, thus suggesting that BBX33 functions downstream of ELONGATED HYPOCOTYL 5 (HY5) and upstream of MYB10 as a component in the environmental sensing pathway, thereby resulting in anthocyanin production in apple (Fig. 3; Bai et al. 2014; Liu et al. 2021).

Liu et al. (2018) assessed the roles of 23 apple *BBX* genes in anthocyanin regulation. Several *BBX* genes, including *Arabidopsis CONSTANS* (*CO)*, *CONSTANS-LIKE1* (*COL1),* and *COL2*, are reported to undergo diurnal patterns of gene expression. Furthermore, several studies investigated diurnal expression patterns in both candidate apple *BBX* genes and anthocyanin structural genes, along with expression profiles of anthocyanin biosynthetic genes and the MYB regulator, *MYB1*/*MYB10* (An et al. 2019, 2020; Liu et al. 2019). A group of BBX proteins has been evaluated for their ability to activate the promoter of *MYB1/MYB10* (Plunkett et al. 2019).

The anthocyanin biosynthesis pathway is influenced by several parameters, including environmental factors (light, temperature, water, and sugar), phytohormones,

transcription factors, as well as epigenetic modifications (Wang and Chen 2021). For instance, it is reported that light negatively regulates anthocyanin biosynthesis in apple primarily by inhibiting expression of *MdBBX37,* as *MdBBX* reduces expression of *MdHY5* by directly targeting its promoter (Liu et al. 2019). In contrast, the transcription factors MdWRKY72 and MdWRKY11 bind to the W-box cis-element of *MdHY5* to activate its regulation of anthocyanin biosynthesis (Liu et al. 2019; Hu et al. 2020). Therefore, it is apparent that BBX proteins function in response to light-induced anthocyanin accumulation and requiring HY5 participation. Furthermore, MdHY5, a bZIP TF and positive regulator of light signaling, can enhance anthocyanin biosynthesis via direct activation of *MdMYB1/10* expression (An et al. 2019). Nevertheless, MdBBX37 hinders binding of MdMYB1 and MdMYB9 to their respective target genes, thereby reducing accumulation of anthocyanin (Ban et al. 2007; Takos et al. 2006; Liu et al. 2021).

Phytohormones such as abscisic acid (ABA), jasmonic acid (JA), auxin, and ethylene are critical for plant and development, and they play important roles in anthocyanin biosynthesis (Wang and Chen 2021). An et al. (2021) have reported that ABI5 promotes ABA-induced anthocyanin biosynthesis by regulating the MYB1-bHLH3 complex in apple. On the other hand, JA induces degradation of JAZ (jasmonate ZIM-domain) proteins in apple via direct interaction of MdJAZ with MdbHLH3, thereby interfering with the recruitment of MdbHLH3 to the promoter of *MdMYB9* resulting in repressed transcription of the MBW complexes which leads to reduction in anthocyanin biosynthesis. As it has been observed that ethylene inhibits anthocyanin biosynthesis in red pears by downregulating expression of R2R3-MYBs (including *PpMYB10* and *PpMYB114*) and LBGs (late biosynthetic genes), and as ethylene signal transduction takes place via ethylene response factors (ERFs) and ethylene-insensitive 3 (EIN3)/EIN3-like (EIL), it has been reported that MdERF1B binds to promoters of *MdMYB9* and *MdMYB11* in regulating accumulation of both anthocyanin and proanthocyanidin (Zhang et al. 2018). Furthermore, elevated auxin levels can inhibit anthocyanin biosynthesis by suppressing structural and regulatory genes, while gibberellic acid (GA) signaling interrupts anthocyanin biosynthesis via DELLA proteins, essential for GA's role in regulating anthocyanin biosynthesis (Liu et al. 2021).

Expression of structural genes during anthocyanin biosynthesis is directly under the control of the MYB-bHLH-WDR complex, and R2R3-MYB TFs are highly involved in the regulatory anthocyanin pathway as these directly regulate expression of related genes, thereby resulting in tissue-specific accumulation of anthocyanin accumulation (Liu et al. 2021). Furthermore, BHLH TFs are critical for the activity of R2R3-MYBs, by either stabilizing the MYB complex or promoting its transcription (Liu et al. 2021). For instance, some MYB TFs, such as MdMYB1, MdMYB9, MdMYB10, and MdMYB114, can promote red color development in apple fruit via their interactions with bHLH3 and WD40 (Ban et al. 2007; Jiang et al. 2021). The MdMYB1 is detected in the skin tissue of apple, and its product participates in photoinduction by activating transcription activities of promoters of *MdDFR* and *MdUFGT*. It has been observed that *MdMYB1* transcript levels are positively correlated with anthocyanin, as well as in accumulation and expression

of structural genes as promoters of both *MdDFR* and *MdUFG* have light-response elements, such as ACGTs and MRE or MRE-like sequences (Hartmann et al. 2005; Takos et al. 2006). As members of the WD40 protein family have 4–10 random WD repeat domains, consisting of 40 amino acid sequences ending in tryptophan (W) and aspartic acid (D), MdTTG1 was the first WD40 protein isolated from apple and found to interact with both MdbHLH3 and MdMYB9 in controlling expression of downstream structural genes (Brueggemann et al. 2010). As it has been observed that some MYB TFs, such as apple MYB16, negatively regulate anthocyanin biosynthesis (Xu et al. 2017; Liu et al. 2021; Li et al. 2022), interactions of such MYBs with bHLH would hinder formation of the MBW complex, as well as competing in their interactions with subunits of bHLH and MYB/bHLH (Liu et al. 2021; Li et al. 2022). As TFs of members of the same family play different roles in the regulation of anthocyanin biosynthesis, functional elucidation of these TFs remains in progress.

In addition to the above TF factors, other regulatory factors such as methylation and demethylation of DNA are also involved in the regulation of anthocyanin biosynthesis. For instance, the DNA methylation inhibitor 5-azacytidine can induce red color pigmentation in apple (Liu et al. 2021). Furthermore, microRNAs can play critical roles in anthocyanin biosynthesis. Zhang et al. (2020) have reported that expression of mdm-miR828 only increases during late red color development in the skin of apple fruit and that this miRNA is involved in a feedback regulatory mechanism associated with anthocyanin accumulation in apple. Moreover, mdm-miR828 is reported to inhibit accumulation of anthocyanin in response to high temperature (Zhang et al. 2020).

5.17 Future Opportunities and Challenges of Apple Nutraceutomics

In the past decade, there here has been a significant interest in enhancing the phytochemical content of apple fruit, particularly of various secondary metabolites involved in the phenylpropanoid, flavonoid, and anthocyanin biosynthesis pathways, as well as of vitamin C and fiber content. With availability of various biochemical analysis and metabolomic platforms and tools, as well as of molecular, genomic, transcriptomic, and bioinformatic tools and platforms, there are ongoing concerted and focused efforts in the evolving field of nutraceutomics. Nutraceuticals have been gaining more considerable interest due to their demonstrated safe, health benefit, and therapeutic effects. Many studies have focused on apple nutraceuticals, primarily of those secondary metabolites involved in the flavonoid biosynthesis pathway such as phloridzin, phloretin, and phenolic acids as they have demonstrated not only high antioxidant defense activities, as well as their critical roles in cell proliferation and gene expression, but they are also more readily bioavailable than those of other phytochemicals, particularly of anthocyanins. Therefore, nutraceuticals offer opportunities for improving human health, preventing chronic diseases, and alleviating various ailments and diseases, such as cancer, diabetes, cardiovascular diseases, asthma, pulmonary

diseases, gastrointestinal disorders, and neurological disorders, among others. The most critical role of these apple nutraceuticals is to protect against oxidative damage, as humans live in a highly oxidative environment and various metabolic processes may result in production of higher levels of oxidants. As oxidative damage accumulates, it becomes more important to protect against oxidative damage early on.

When compared to other fruits consumed in the USA, apples demonstrate the second highest level of antioxidant activity, after cranberry (Boyer and Liu 2003–2004). Apples rank second for total content of phenolic compounds. When compared to other fruits, apples have the highest proportion of free phenolics; thus, these phenolics are likely to be more bioavailable, for eventual absorption into the bloodstream. As noted above, several of the antioxidant compounds in apples have been well investigated such as quercetin, catechin, epicatechin, procyanidin, chlorogenic acid, and phloridzin.

Nutraceutomics offers new opportunities for pursuing enhancement of nutraceutical content of apple fruit in advanced breeding lines via integration of metabolome profiles, antioxidant activity, environment, and gene expression during pursuit of a genomic selection scheme(s) in an apple breeding program. Therefore, it is important that considerations of environmental influences, such as temperature and irradiation, as well as of fruit developmental factors are accounted for in gene expression analysis and in metabolic responses of these selections, as these present challenges in achieving rapid advances in nutraceutomics. Moreover, epigenetic modifications via DNA methylation can play important roles in structural and regulatory gene expression of "nutraceutical" profile(s).

However, the key factor in continued successful advances in the field of apple nutraceutomics is to develop a set of priorities of those metabolites that are likely to be highly expressed in apple fruits under varying environmental conditions, particularly in light of current conditions of climate change, and also to demonstrate high efficacies of bioavailability. This should be followed by genetic enhancement and breeding efforts via various breeding schemes and strategies. These strategies could be pursued using either marker-assisted selection, genomic selection, rapid-cycling breeding, or gene editing, among others.

6 Conclusions

Plant secondary metabolites in apple fruit have been documented to be involved in alleviating various human health diseases and conditions, and therefore there has been increased interest in enhancing levels of these nutraceutical compounds in both skin and flesh tissues of apple fruits, as well as their bioavailability and antioxidant activities (Busatto et al. 2019; Nezbedova et al. 2021; Koh et al. 2019). Various studies have assessed contents of these nutraceutical compounds in different cultivars, genotypes, and accessions, and although contents of these compounds are higher in wild versus cultivated apples, there are various opportunities to enhance accumulation of these compounds in the cultivated apple (Busatto et al. 2019). Genetic studies have been undertaken to identify and map genes and

QTLs associated with these compounds, particularly of phytochemical compounds involved in the phenylpropanoid pathway, including polyphenolic compounds consisting of phenolic acids and flavonoids (McClure et al. 2019; Brizzolara et al. 2021; Kim et al. 2020a). As marker-assisted selection efforts have become integrated in modern breeding programs, availability of molecular markers associated with various polyphenolic compounds has become highly critical in selecting genotypes with enhanced levels of these different compounds (Teh et al. 2021). Furthermore, owing to the availability of tools of apple genomics, transcriptomics, metabolomics, and other omics technologies, the field of apple nutraceutomics is a new frontier that is gaining stronger interest by apple geneticists, nutritionists, and health professionals (Korban 2021; Kidoń and Grabowska 2021; Li et al. 2020; Nasir et al. 2020).

References

Alam MN, Almoyad M, Huq F (2018) Polyphenols in colorectal cancer: current state of knowledge including clinical trials and molecular mechanism of action. Biomed Res Int 2018:4154185

Albert NW, Davies KM, Lewis DH, Zhang H, Montefiori M, Brendolise C, Boase MR, Ngo H, Jameson PE, Schwinn KE (2014) A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. Plant Cell 26:962–980. https://doi.org/10.1105/tpc.113.122069

Alberti A, Zielinski A, Couto M, Judacewski P, Igarashi-Mafra L, Nogueira A (2017) Distribution of phenolic compounds and antioxidant capacity in apples tissues during ripening. J Food Sci Technol 54:1511–1518

Allan A, Hellens R, Laing W (2008) MYB transcription factors that colour our fruit. Trends Plant Sci 13:99–102

Almanza-Aguilera E, Ceballos-Sánchez D, Achaintre D, Rothwell JA, Laouali N, Severi G, Katzke V, Johnson T, Schulze MB, Palli D, Gargano G, de Magistris MS, Tumino R, Sacerdote C, Scalbert A, Zamora-Ros R (2021) Urinary concentrations of (+)-catechin and (−)-epicatechin as biomarkers of dietary intake of flavan-3-ols in the European prospective investigation into cancer and nutrition (EPIC) study. Nutrients 13(11):4157. https://doi.org/10.3390/nu13114157

Almeida DP, Gião MS, Pintado M, Gomes MH (2017) Bioactive phytochemicals in apple cultivars from the Portuguese protected geographical indication "Maçã de Alcobaça": basis for market segmentation. Int J Food Prop 20:2206–2214

Amyotte B, Bowen AJ, Banks T, Rajcan I, Somers DJ (2017) Mapping the sensory perception of apple using descriptive sensory evaluation in a genome wide association study. PLoS One 12:e0171710

An XH, Tian Y, Chen KQ, Wang XF, Hao YJ (2012) The apple WD40 protein MdTTG1 interacts with bHLH but not MYB proteins to regulate anthocyanin accumulation. J Plant Physiol 169:710–717

An J-P, Wang X-F, Espley R, Lin-Wang K, Bi S-Q, You C-X, Hao Y-J (2019) An apple B-box protein MdBBX37 modulates anthocyanin biosynthesis and hypocotyl elongation synergistically with MdMYBs and MdHY5. Plant Cell Physiol 61:130–143. https://doi.org/10.1093/pcp/pcz185

An J-P, Liu Y-J, Zhang X-W, Bi S-Q, Wang X-F, You C-X, Hao Y-J (2020) Dynamic regulation of anthocyanin biosynthesis at different light intensities by the BT2-TCP46-MYB1 module in apple. J Exp Bot 71:3094–3109. https://doi.org/10.1093/jxb/eraa056

An JP, Zhang XW, Liu YJ, Wang XF, You CX, Hao YJ (2021) ABI5 regulates ABA-induced anthocyanin biosynthesis by modulating the MYB1-bHLH3 complex in apple. J Exp Bot 72:1460–1472. https://doi.org/10.1093/jxb/eraa525

Aprikian O, Duclos V, Guyot S, Besson C, Manach C, Bernalier A, Morand C, Rémésy C, Demigné C (2003) Apple pectin and a polyphenol-rich apple concentrate are more effective together than separately on cecal fermentations and plasma lipids in rats. J Nutr 133:1860–1865

Awad M, de Jager A (2002a) Relationships between fruit nutrients and concentrations of flavonoids and chlorogenic acid in Elstar apple peels. Sci Hortic 92:265–276

Awad M, de Jager A (2002b) Formation of flavonoids, especially anthocyanin and chlorogenic acid in Jonagold apple peels: influences of growth regulators and fruit maturity. Sci Hortic 93:257–266

Awad M, de Jager A, van der Plas L, van der Krol A (2001a) Flavonoid and chlorogenic acid changes in peels of Elstar and Jonagold apples during development and ripening. Sci Hortic 90:69–83

Awad M, Wagenmakers P, de Jager A (2001b) Effects of light on flavonoid and chlorogenic acid levels in the peels of Jonagold apples. Sci Hortic 88:289–298

Bai S, Saito T, Honda C, Hatsuyama Y, Ito A, Moriguchi T (2014) An apple B-box protein, MdCOL11, is involved in UV-B- and temperature-induced anthocyanin biosynthesis. Planta 240:1051–1062. https://doi.org/10.1007/s00425-014-2129-8

Ban Y, Honda C, Hatsuyama Y, Igarashi M, Bessho H, Moriguchi T (2007) Isolation and functional analysis of a MYB transcription factor gene that is a key regulator for the development of red coloration in apple skin. Plant Cell Physiol 48:958–970. https://doi.org/10.1093/pcp/pcm066

Bao M, Shen J, Jia Y, Li F, Ma W, Shen H, Shen L, Lin X, Zhang L, Dong X (2013) Apple polyphenol protects against cigarette smoke-induced acute lung injury. Nutrition 29:235–243. https://doi.org/10.1016/j.nut.2012.04.008

Bars-Cortina D, Martínez-Bardají A, Macià A, Motilva MJ, Piñol-Felis C (2020) Consumption evaluation of one apple flesh a day in the initial phases prior to adenoma/adenocarcinoma in an azoxymethane rat colon carcinogenesis model. J Nutr Biochem 83:108418

Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L (2004) TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. Plant J 39:366–380

Bazzano LA, Serdula MK, Liu S (2003) Dietary intake of fruits and vegetables and risk of cardiovascular disease. Curr Atheroscler Rep 5:492–499

Begić-Akagić A, Spaho N, Oručević S, Drkenda P, Kurtović M, Gaši F, Kopjar M, Piližota V (2011) Influence of cultivar, storage time, and processing on the phenol content of cloudy apple juice. Croat J Food Sci Technol 3(2):1–8

Blesso CN (2019) Dietary anthocyanins and human health. Nutrients 11:2107

Boeing H, Bechthold A, Bub A, Ellinger S, Haller D, Kroke A, Leschik-Bonnet E, Müller MJ, Oberritter H, Schulze M, Stehle P, Watzl B (2012) Critical review: vegetables and fruit in the prevention of chronic diseases. Eur J Nutr 51:637–663. https://doi.org/10.1007/s00394-012-0380-y

Bondonno NP, Bondonno CP, Ward NC, Hodgson JM, Croft KD (2017) The cardiovascular health benefits of apples: whole fruit vs. isolated compounds. Trends Food Sci Tech 69(Part B):243–256

Bondonno NP, Davey RJ, Murray K, Radavelli-Bagatini S, Bondonno CP, Blekkenhorst LC, Sim M, Magliano DJ, Daly RM, Shaw JE, Lewis JR, Hodgson JM (2021) Associations between fruit intake and risk of diabetes in the AusDiab Cohort. J Clin Endocrinol Metab 106(10): e4097–e4108. https://doi.org/10.1210/clinem/dgab335

Borevitz JO, Xia YJ, Blount J, Dixon RA, Lamb C (2000) Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell 12:2383–2394

Boudichevskaia A, Kumar G, Sharma Y, Kapoor R, Singh AK (2020) Challenges and strategies for developing climate-smart apple varieties through genomic approaches. In: Kole C (ed) Genomic designing of climate-smart fruit crops. Springer, Cham, pp 23–71

Bowler RP (2004) Oxidative stress in the pathogenesis of asthma. Curr Allergy Asthma Rep 4:116–122. https://doi.org/10.1007/s11882-004-0056-7

Boyer J, Liu RH (2003–2004) Antioxidants of apples. N Y Fruit Q 11(4):11–15

Boyer J, Liu RH (2004) Apple phytochemicals and their health benefits. Nutr J 3:5. https://doi.org/10.1186/1475-2891-3-5

Brazier Y (2019) What to know about apples. Medical News Today Newsletter. https://www.medicalnewstoday.com/articles/267290

Brglez Mojzer E, Knez Hrnčič M, Škerget M, Knez Ž, Bren U (2016) Polyphenols: extraction methods, antioxidative action, bioavailability and anticarcinogenic effects. Molecules 21:1–38

Briguglio G, Costa C, Pollicino M, Giambò F, Catania S, Fenga C (2020) Polyphenols in cancer prevention: new insights. Int J Funct Nutr 1:9. https://doi.org/10.3892/ijfn.2020.9

Brizzolara S, Tenori L, Korban SS (2021) Metabolomic approaches for apple fruit quality improvement. In: Korban SS (ed) The apple genome. Springer, Cham, pp 311–340

Brueggemann J, Weisshaar B, Sagasser M (2010) A WD40-repeat gene from *Malus* x *domestica* is a functional homologue of *Arabidopsis thaliana TRANSPARENT TESTA GLABRA1*. Plant Cell Rep 29:285–294. https://doi.org/10.1007/s00299-010-0821-0

Busatto N, Matsumoto D, Tadiello A, Vrhovsek U, Costa F (2019) Multifaceted analyses disclose the role of fruit size and skin-russeting in the accumulation pattern of phenolic compounds in apple. PLoS One 14(7):e0219354. https://doi.org/10.1371/journal.pone.0219354

Butkeviciute A, Viskelis J, Liaudanskas M, Viskelis P, Bobinas C, Janulis V (2021) Variation of triterpenes in apples stored in a controlled atmosphere. Molecules 26:3639. https://doi.org/10.3390/molecules26123639

Carbone K, Giannini B, Picchi V, Scalzo RL, Cecchini F (2011) Phenolic composition and free radical scavenging activity of different apple varieties in relation on the cultivar, tissue type and storage. Food Chem 127:493–500. https://doi.org/10.1016/j.foodchem.2011.01.030

Celton J-M, Bianco L, Linsmith G, Balzerque S, Troggio M (2021) The apple genome and epigenome. In: Korban SS (ed) The apple genome. Springer, Cham, pp 169–187

Chagné D, Carlisle CM, Blond C, Volz RK, Whitworth CJ, Oraguzie NC, Crowhurst RN, Allan AC, Espley RV, Hellens RP, Gardiner SE (2007) Mapping a candidate gene (*MdMYB10*) for red flesh and foliage colour in apple. BMC Genomics 8:212

Chagné D, Krieger C, Rassam M, Sullivan M, Fraser J, André C, Pindo M, Troggio M, Gardiner SE, Henry RA, Allan AC, McGhie TK, Laing WA (2012) QTL and candidate gene mapping for polyphenolic composition in apple fruit. BMC Plant Biol 12:12. https://doi.org/10.1186/1471-2229-12-12

Chagné D, Lin-Wang K, Espley RV, Volz RK, How NM, Rouse S, Brendolise C, Carlisle CM, Kumar S, De Silva N, Micheletti D, McGhie T, Crowhurst RN, Storey RD, Velasco R, Hellens RP, Gardiner SE, Allan AC (2013) An ancient duplication of apple MYB transcription factors is responsible for novel red fruit-flesh phenotypes. Plant Physiol 161:225–239. https://doi.org/10.1104/pp.112.206771

Chagné D, Kirk C, How N, Whitworth C, Fontic C, Reig G, Sawyer G, Rouse S, Poles L, Gardiner SE, Kumar S, Espley R, Volz RK, Troggio M, Iglesias I (2016) A functional genetic marker for apple red skin coloration across different environments. Tree Genet Genomes 12:67. https://doi.org/10.1007/s11295-016-1025-8

Chan A, Shea T (2009) Dietary supplementation with apple juice decreases endogenous amyloid-beta levels in murine brain. J Alzheimers Dis 16:167–171

Chen C, Lim Y (2018) Receptor to attenuate ligand-induced lipogenesis. J Agric Food Chem 66:10964–10976

Chen M, Gowd V, Wang M, Chen F, Cheng K-W (2020) The apple dihydrochalcone phloretin suppresses growth and improves chemosensitivity of breast cancer cells via inhibition of cytoprotective autophagy. Food Funct 12:177–190

Choi BY (2019) Biochemical basis of anti-cancer-effects of phloretin – a natural dihydrochalcone. Molecules 24(2):278. https://doi.org/10.3390/molecules24020278

Clifford MN, Jaganath IB, Ludwig IA, Crozier A (2017) Chlorogenic acids and the acyl-quinic acids: discovery, biosynthesis, bioavailability and bioactivity. Nat Prod Rep 34:1391–1421

Cosme P, Rodríguez AB, Espino J, Garrido M (2020) Plant phenolics: bioavailability as a key determinant of their potential health-promoting applications. Antioxidants 9:1263

Cuthbertson D, Andrews PK, Reganold JP, Davies NM, Lange BM (2012) Utility of metabolomics toward assessing the metabolic basis of quality traits in apple fruit with an emphasis on antioxidants. J Agric Food Chem 60:8552–8560. https://doi.org/10.1021/jf3031088

Dashbaldan S, Pączkowski C, Szakiel A (2020) Variations in triterpenoid deposition in cuticular waxes during development and maturation of selected fruits of Rosaceae family. Int J Mol Sci 21:9762

Daveri E, Adamo AM, Alfine E, Zhu W, Oteiza PI (2020) Hexameric procyanidins inhibit colorectal cancer cell growth through both redox and non-redox regulation of the epidermal growth factor signaling pathway. Redox Biol 38:101830

Davey MW, Kenis K, Keulemans J (2006) Genetic control of fruit vitamin C contents. Plant Physiol 142:343–351

Davies KM, Jibran R, Zhou Y, Albert NW, Brummell DA, Jordan BR, Bowman JL, Schwinn KE (2020) The evolution of flavonoid biosynthesis: a bryophyte perspective. Front Plant Sci 11:7. https://doi.org/10.3389/fpls.2020.00007

de la Iglesias R, Milagro FI, Boqué N, Martinez JA (2010) Healthy properties of proanthocyanidins. Biofactors 36:159–168. https://doi.org/10.1002/biof.79

De Paepe D, Valkenborg D, Noten B, Servaes K, Diels L, De Loose M, Van Droogenbroeck B, Voorspoels S (2015) Variability of the phenolic profiles in the fruits from old, recent and new apple cultivars cultivated in Belgium. Metabolomics 11:739–752

Devereux G, Seaton A (2005) Diet as a risk factor for atopy and asthma. J Allergy Clin Immunol 115:1109–1117

Di Lorenzo C, Colombo F, Biella S, Stockley C, Restani P (2021) Polyphenols and human health: the role of bioavailability. Nutrients 13:273. https://doi.org/10.3390/nu13010273

Dickson EE, Arumuganathan K, Kresovich S, Doyle JJ (1992) Nuclear DNA content variation within the Rosaceae. Am J Bot 79:1081–1086. https://doi.org/10.2307/2444917

Dunemann F, Ulrich D, Boudichevskaia A, Grafe C, Weber WE (2009) QTL mapping of aroma compounds analysed by headspace solid-phase microextraction gas chromatography in the apple progeny "Discovery" × "Prima". Mol Breed 23:501–521

Efimtseva EA, Chelpanova TI (2020) Apples as a source of soluble and insoluble dietary fibers: effect of dietary fibers on appetite. Hum Physiol 46:224–234

Escarpa A, Gonzalez M (1998) High-performance liquid chromatography with diode-array detection for the performance of phenolic compounds in peel and pulp from different apple varieties. J Chromatography A 823:331–337. https://doi.org/10.1016/S0021-9673(98)00294-5

Espley RV, Hellens RP, Putterill J, Stevenson DE, Kutty-Amma S, Allan AC (2007) Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. Plant J 49:414–427. https://doi.org/10.1111/j.1365-313X.2006.02964.x

Espley RV, Brendolise C, Chagne D, Kutty-Amma S, Green S, Volz R, Putterill J, Schouten HJ, Gardiner SE, Hellens RP, Allan AC (2009) Multiple repeats of a promoter segment causes transcription factor autoregulation in red apples. Plant Cell 21:168–183. https://doi.org/10.1105/tpc.108.059329

Espley RV, Bovy A, Bava C, Jaeger SR, Tomes S, Norling C, Crawford J, Rowan D, McGhie TK, Brendolise C, Putterill J, Schouten HJ, Hellens RP, Allan AC (2013) Analysis of genetically modified red-fleshed apples reveals effects on growth and consumer attributes. Plant Biotechnol J 11:408–419. https://doi.org/10.1111/pbi.12017

Fabiani R, Minelli L, Rosignoli P (2016) Apple intake and cancer risk: a systematic review and meta-analysis of observational studies. Public Health Nutr 19:2603–2617

Fang T, Zhen Q, Liao L, Owiti A, Zhao L, Korban SS, Han Y (2017) Variation of ascorbic acid concentration in fruits of cultivated and wild apples. Food Chem 225:132–137. https://doi.org/10.1016/j.foodchem.2017.01.014

Feng S, Yi J, Li X, Wu X, Zhao Y, Ma Y, Bi J (2021) Systematic review of phenolic compounds in apple fruits: compositions, distribution, absorption, metabolism, and processing stability. J Agric Food Chem 69:7–27. https://doi.org/10.1021/acs.jafc.0c05481

Fernandes I, de Freitas V, Mateus N (2014) Anthocyanins and human health: how gastric absorption may influence acute human physiology. Nutr Aging 2:1–14. https://doi.org/10.3233/NUA-130030

Fernando W, Rupasinghe HV, Hoskin DW (2019) Dietary phytochemicals with anti-oxidant and pro-oxidant activities: a double-edged sword in relation to adjuvant chemotherapy and radiotherapy? Cancer Lett 452:168–177

Food and Agricultural Organization – FAOSTAT (2022) http://faostat.fao.org/. Accessed 20 Jan 2022

Francini A, Sebastiani L (2013) Phenolic compounds in apple (*Malus x domestica* Borkh.): compounds characterization and stability during postharvest and after processing. Antioxidants 2:181–193. https://doi.org/10.3390/antiox2030181

Fukumitsu S, Villareal MO, Fujitsuka T, Aida K (2016) Anti-inflammatory and anti-arthritic effects of pentacyclic triterpenoids maslinic acid through NF-κB inactivation. Mol Nutr Food Res 60: 399–409

Furtado NAJC, Pirson L, Edelberg H, Miranda LM, Loira-Pastoriza C, Preat V, Larondelle Y, André CM (2017) Pentacyclic triterpene bioavailability: an overview of in vitro and in vivo studies. Molecules 22:400

Galvis-Sanchez AC, Rocha A (2016) Bioactive compounds of apples and pears as health promoters. In: Silva LR, Silva B (eds) Natural bioactive compounds from fruits and vegetables as health promoters. Bentham Science Publishers, Sharjah, pp 98–109

Gayer BA, Avendano EE, Edelson E, Nirmala N, Johnson EJ, Raman G (2019) Effects of intake of apples, pears, or their products on cardiometabolic risk factors and clinical outcomes: a systematic review and meta-analysis. Curr Dev Nutr 3:nzz109

Guo XF, Yang B, Tang J, Jiang JJ, Li D (2017) Apple and pear consumption and type 2 diabetes mellitus risk: a meta-analysis of prospective cohort studies. Food Funct 8:927–934

Guo H, Ding H, Tang X, Liang M, Li S, Zhang J, Cao J (2021) Quercetin induces pro-apoptotic autophagy via SIRT1/AMPK signaling pathway in human lung cancer cell lines A549 and H1299 in vitro. Thorac Cancer 12:1415–1422

Guyenet SJ (2019) Impact of whole, fresh fruit consumption on energy intake and adiposity: a systematic review. Front Nutr 6:66

Guyot S, Le Bourvellec C, Marnet N, Drilleau JF (2002) Procyanidins are the most abundant polyphenols in dessert apples at maturity. LWT Food Sci Technol 35:289–291

Guyot S, Marnet N, Sanoner P, Drilleau JF (2003) Variability of the polyphenolic composition of cider apple (*Malus domestica*) fruits and juices. J Agric Food Chem 51:6240–6247

Hammerstone J, Lazarus S, Schmitz H (2000) Procyanidin content and variation in some commonly consumed foods. J Nutr 130:2086S–2092S

Han Y, Korban SS (2021) Genetic and physical mapping of the apple genome. In: Korban SS (ed) The apple genome. Springer, Cham, pp 131–168

Han Y, Gasic K, Korban SS (2007) Multiple-copy cluster-type organization and evolution of genes encoding O-methyltransferases in apple. Genetics 176:2625–2635

Han Y, Vimolmangkang S, Soria-Guerra R, Rosales-Mendoza S, Zheng D, Lygin AV, Korban SS (2010) Ectopic expression of apple *F3′H* genes contribute to the accumulation of anthocyanin in the Arabidopsis tt7 mutant grown under N-deficient conditions. Plant Physiol 153:806–820

Han Y, Vimolmangkang S, Soria-Guerra RE, Korban SS (2012) Introduction of apple ANR genes into tobacco inhibits expression of both CHI and DFR genes in flowers, leading to loss of anthocyanin. J Exp Bot 63:2437–2447. https://doi.org/10.1093/jxb/err415

Han M, Li G, Liu X, Li A, Mao P, Liu P, Li H (2019) Phenolic profile, antioxidant activity and antiproliferative activity of crabapple fruits. Hortic Plant J 5:155–163. https://doi.org/10.1016/j.hpj.2019.01.003

Hanhan D (2022) Hallmarks of cancer: new dimensions. Cancer Discov 12:31–46. https://doi.org/10.1158/2159-8290.CD-21-105

Hartmann U, Sagasser M, Mehrtens F, Stracke R, Weisshaar B (2005) Differential combinatorial interactions of cis-acting elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes. Plant Mol Biol 57:155–171. https://doi.org/10.1007/s11103-004-6910-0

Henry-Kirk RA, McGhie TK, Andre CM, Hellens RP, Allan AC (2012) Transcriptional analysis of apple fruit proanthocyanidin biosynthesis. J Exp Bot 63:5437–5450. https://doi.org/10.1093/jxb/ers193

Hichri I, Heppel SC, Pillet J, Léon C, Czemmel S, Delrot S, Lauvergeat V, Bogs J (2010) The basic helix-loop-helix transcription factor MYC1 is involved in the regulation of the flavonoid biosynthesis pathway in grapevine. Mol Plant 3:509–523. https://doi.org/10.1093/mp/ssp118

Ho KK, Ferruzzi MG, Wightman JD (2020) Potential health benefits of (poly) phenols derived from fruit and 100% fruit juice. Nutr Rev 78:145–174. https://doi.org/10.1093/nutrit/nuz041

Hodgson JM, Prince RL, Woodman RJ, Bondonno CP, Ivey KL, Bondonno N, Rimm EB, Ward NC, Croft KD, Lewis JR (2016) Apple intake is inversely associated with all-cause and disease-specific mortality in elderly women. Br J Nutr 115(5):860–867

Hoffmann L, Besseau S, Geoffroy P, Ritzenthaler C, Meyer D, Lapierre C, Pollet B, Legrand M (2004) Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyl-transferase affects phenylpropanoid biosynthesis. Plant Cell 16:1446–1465. https://doi.org/10.1105/tpc.020297

Honda C, Moriya S (2018) Anthocyanin biosynthesis in apple fruit. Hortic J 87:305–314

Hosseini B, Berthon BS, Wark P, Wood LG (2017) Effects of fruit and vegetable consumption on risk of asthma, wheezing and immune responses: a systematic review and meta-analysis. Nutrients 9(4):341. https://doi.org/10.3390/nu9040341

Hou N, Liu N, Han J, Yan Y, Li J (2017) Chlorogenic acid induces reactive oxygen species generation and inhibits the viability of human colon cancer cells. Anti-Cancer Drugs 28:59–65

Hu J, Fang H, Wang J, Yue X, Su M, Mao Z, Zou Q, Jiang H, Guo Z, Yu L, Feng T, Lu L, Peng Z, Zhang Z, Wang N, Chen X (2020) Ultraviolet B-induced *MdWRKY72* expression promotes anthocyanin synthesis in apple. Plant Sci 292:110377. https://doi.org/10.1016/j.plantsci.2019.110377

Hyson DA (2011) A comprehensive review of apples and apple components and their relationship to human health. Adv Nutr 2(5):408–420. https://doi.org/10.3945/an.111.000513

Hyun TK, Jang KI (2016) Apple as a source of dietary phytonutrients: an update on the potential health benefits of apple. EXCLI J 15:565

Ichwan M, Walker TL, Nicola Z, Ludwig-Müller J, Böttcher C, Overall RW, Adusumilli VS, Bulut M, Sykes AM, Hübner N, Ramirez-Rodriguez G, Ortiz-López L, Lugo-Hernández EA, Kempermann G (2021) Apple peel and flesh contain pro-neurogenic compounds. Stem Cell Rep 16(3):548–565. https://doi.org/10.1016/j.stemcr.2021.01.005

Imeh U, Khokhar S (2002) Distribution of conjugated and free phenols in fruits: antioxidant activity and cultivar variations. J Agric Food Chem 50:6301–6306

Jakopic J, Stampar F, Veberic R (2009) The influence of exposure to light on the phenolic content of "Fuji" apple. Sci Hortic 123:234–239

Jia L, Huang S, Yin X, Zan Y, Guo Y, Han L (2018) Quercetin suppresses the mobility of breast cancer by suppressing glycolysis through Akt-mTOR pathway mediated autophagy induction. Life Sci 208:123–130

Jiang S, Sun Q, Zhang T, Liu W, Wang N, Chen X (2021) *MdMYB114* regulates anthocyanin biosynthesis and functions downstream of MdbZIP4-like in apple fruit. J Plant Physiol 257:153353. https://doi.org/10.1016/j.jplph.2020.153353

Jugdé H, Nguy D, Moller I, Cooney JM, Atkinson RG (2008) Isolation and characterization of a novel glycosyltransferase that converts phloretin to phlorizin, a potent antioxidant in apple. FEBS J 275:3804–3814

Kaeswurm JAH, Burandt MR, Mayer PS, Straub LV, Buchweitz M (2022) Bioaccessibility of apple polyphenols from peel and flesh during oral digestion. J Agric Food Chem. https://doi.org/10.1021/acs.jafc.1c08130

Kalinowska M, Bielawska A, Lewandowska-Siwkiewicz H, Priebe W, Lewandowski W (2014) Apples: content of phenolic compounds vs. variety, part of apple and cultivation model, extraction of phenolic compounds, biological properties. Plant Physiol Biochem 84:169–188

Kalinowska M, Gryko K, Wróblewska AM, Jabłońska-Trypuć A, Karpowicz D (2020) Phenolic content, chemical composition and anti-/pro-oxidant activity of Gold Milenium and Papierowka apple peel extracts. Sci Rep 10:14951

Kawser Hossain M, Abdal Dayem A, Han J, Yin Y, Kim K, Saha SK, Yang G-M, Choi HY, Cho S-G (2016) Molecular mechanisms of the anti-obesity and anti-diabetic properties of flavonoids. Int J Mol Sci 17:569. https://doi.org/10.3390/ijms17040569

Khan A, Korban SS (2022) Breeding and genetics of disease resistance in temperate fruit trees: challenges and new opportunities. Theor Appl Genet 135:3961–3985. https://doi.org/10.1007/s00122-022-04093-0

Khan SA, Chibon PY, de Vos RC, Schipper BA, Walraven E, Beekwilder J, van Dijk T, Finkers R, Visser RG, van de Weg EW, Bovy A, Cestaro A, Velasco R, Jacobsen E, Schouten HJ (2012a) Genetic analysis of metabolites in apple fruits indicates an mQTL hotspot for phenolic compounds on linkage group 16. J Exp Bot 63(8):2895–2908. https://doi.org/10.1093/jxb/err464

Khan SA, Schaart JG, Beekwilder J, Allan AC, Tikunov YM, Jacobsen E, Schouten HJ (2012b) ThemQTL hotspot on linkage group 16 for phenolic compounds in apple fruits is probably the result of a leucoanthocyanidin reductase gene at that locus. BMC Res Notes 5:618

Khan A, Gutierrez B, Chao T, Singh J (2021) Origin of the domesticated apple. In: Korban SS (ed) The apple genome. Springer, Cham, pp 383–394

Kidoń M, Grabowska J (2021) Bioactive compounds, antioxidant activity, and sensory qualities of red-fleshed apples dried by different methods. LWT 136:110302. https://doi.org/10.1016/j.lwt.2020.110302

Kim HS, Kim B-G, Sung S, Kim M, Mok H, Chong Y, Ahn J-H (2013) Engineering flavonoid glycosyltransferases for enhanced catalytic efficiency and extended sugar-donor selectivity. Planta 238:683–693

Kim I, Ku KH, Jeong MC, Kim SS, Mitchell AE, Lee J (2019) A comparison of the chemical composition and antioxidant activity of several new early- to mid-season apple cultivars for a warmer climate with traditional cultivars. J Sci Food Agric 99:4712–4724

Kim I, Ku K-H, Jeong M-C, Kwon S-I, Lee J (2020a) Metabolite profiling and antioxidant activity of 10 new early- to mid-season apple cultivars and 14 traditional cultivars. Antioxidants 9:443. https://doi.org/10.3390/antiox9050443

Kim U, Kim CY, Lee JM, Oh H, Ryu B, Kim J, Park JH (2020b) Phloretin inhibits the human prostate cancer cells through the generation of reactive oxygen species. Pathol Oncol Res 26:977–984

Koh Y-C, Ho C-T, Pan M-H (2019) Recent advances in cancer chemoprevention with phytochemicals. J Food Drug Anal 28:14–37

Korban SS (2021) Future prospects for "omics" and other technologies for genetic improvement of apple. In: Korban SS (ed) The apple genome. Springer, Cham, pp 395–412

Koutsos A, Tuohy KM, Lovegrove JA (2015) Apples and cardiovascular health – is the gut microbiota a core consideration? Nutrients 7:3959–3998

Koutsos A, Lima M, Conterno L, Gasperotti M, Bianchi M, Fava F, Vrhovsek U, Lovegrove JA, Tuohy KM (2017) Effects of commercial apple varieties on human gut microbiota composition and metabolic output using an in vitro colonic model. Nutrients 9:533

Koutsos A, Riccadonna S, Ulaszewska MM, Franceschi P, Trošt K, Galvin A, Braune T, Fava F, Perenzoni D, Mattivi F, Tuohy KM, Lovegrove JA (2020) Two apples a day lower serum cholesterol and improve cardiometabolic biomarkers in mildly hypercholesterolemic adults: a randomized, controlled, crossover trial. Am J Clin Nutr 111:307–318. https://doi.org/10.1093/ajcn/nqz282

Kschonsek J, Wolfram T, Stöckl A, Böhm V (2018) Polyphenolic compounds analysis of old and new apple cultivars and contribution of polyphenolic profile to the in vitro antioxidant capacity. Antioxidants 7:20

Łata B (2007) Relationship between apple peel and the whole fruit antioxidant content: year and cultivar variation. J Agric Food Chem 55:663–671

Łata B, Trampczynska A, Paczesna J (2009) Cultivar variation in apple peel and whole fruit phenolic composition. Sci Hortic 121:176–181
Leng E, Xiao Y, Mo Z, Li Y, Zhang Y, Deng X, Li W (2018) Synergistic effect of phytochemicals on cholesterol metabolism and lipid accumulation in HepG2 cells. BMC Complement Altern Med 18:1–10
Li H, Huang CH, Ma H (2019) Whole-genome duplications in pear and apple. In: Korban SS (ed) The pear genome. Springer, Cham, pp 279–299
Li CX, Zhao XH, Zuo WF, Zhang TL, Zhang ZY, Chen XS (2020) Phytochemical profiles, antioxidant, and antiproliferative activities of four red-fleshed apple varieties in China. J Food Sci 85:718–726
Li C, Yu W, Xu J, Lu X, Liu Y (2022) Anthocyanin biosynthesis induced by MYB transcription factors in plants. Int J Mol Sci 23(19):11701. https://doi.org/10.3390/ijms231911701
Liao L, Vimolmangkang S, Wei G, Zhou H, Korban SS, Han Y (2015) Molecular characterization of genes encoding leucoanthocyanidin reductase involved in proanthocyanidin biosynthesis in apple. Front Plant Sci 6:243. https://doi.org/10.3389/fpls.2015.00243
Lin S-T, Tu S-H, Yang P-S, Hsu S-P, Lee WH, Ho C-T, Wu C-H, Lai Y-H, Chen M-Y, Chen L-C (2016) Apple polyphenol phloretin inhibits colorectal cancer cell growth via inhibition of the type 2 glucose transporter and activation of p53-mediated signaling. J Agric Food Chem 64: 6826–6837
Lin-Wang K, Bolitho K, Grafton K, Kortstee A, Karunairetnam S, McGhie TK, Espley RV, Hellens RP, Allan AC (2010) An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. BMC Plant Biol 10:50
Lin-Wang K, Micheletti D, Palmer J, Volz R, Lozano L, Espley R, Hellens RP, Chagnè D, Rowan DD, Troggio M, Iglesias I, Allan AC (2011) High temperature reduces apple fruit colour via modulation of the anthocyanin regulatory complex. Plant Cell Environ 34:1176–1190. https://doi.org/10.1111/j.1365-3040.2011.02316.x
Liu X, Li R, Dai Y, Chen X, Wang X (2018) Genome-wide identification and expression analysis of the *B-box* gene family in the apple (*Malus domestica* Borkh.) genome. Mol Gen Genomics 293: 303–315. https://doi.org/10.1007/s00438-017-1386-1
Liu W, Wang Y, Yu L, Jiang H, Guo Z, Xu H, Jiang S, Fang H, Zhang J, Su M, Zhang Z, Chen X, Chen X, Wang N (2019) MdWRKY11 participates in anthocyanin accumulation in red-fleshed apples by affecting MYB transcription factors and the photoresponse factor MdHY5. J Agric Food Chem 67:8783–8793. https://doi.org/10.1021/acs.jafc.9b02920
Liu H, Liu Z, Wu Y, Zheng L, Zhang G (2021) Regulatory mechanisms of anthocyanin biosynthesis in apple and pear. Int J Mol Sci 22(16):8441. https://doi.org/10.3390/ijms22168441
Luo J, Hu YL, Wang H (2017) Ursolic acid inhibits breast cancer growth by inhibiting proliferation, inducing autophagy and apoptosis, and suppressing inflammatory responses via the PI3K/AKT and NF-κB signaling pathways in vitro. Exp Ther Med 14:3623–3631
Maeda H, Dudareva N (2012) The shikimate pathway and aromatic amino acid biosynthesis in plants. Annu Rev Plant Biol 63:73–105. https://doi.org/10.1146/annurev-arplant-042811-105439
Maffei F, Tarozzi A, Carbone F, Marchesi A, Hrelia S, Angeloni C, Forti G, Hreliaa P (2007) Relevance of apple consumption for protection against oxidative damage induced by hydrogen peroxide in human lymphocytes. Br J Nutr 97:921–927
Mahmud LSM, Tisha A, Sagor AT (2018) A comprehensive review on effective role of apple polyphenols in the treatment of obesity, diabetes, and liver dysfunctions with some possible molecular mechanisms. Oxid Antioxid Med Sci 7:9–27
Marín L, Miguélez EM, Villar CJ, Lombó F (2015) Bioavailability of dietary polyphenols and gut microbiota metabolism: antimicrobial properties. Biomed Res Int 2015:905215
Martínez-Rodríguez OP, Thompson-Bonilla MDR, Jaramillo-Flores ME (2020) Association between obesity and breast cancer: molecular bases and the effect of flavonoids in signaling pathways. Crit Rev Food Sci Nutr 60:3770–3792
Masuda I, Koike M, Nakashima S, Mizutan Y, Ozawa Y, Watanabe K, Sawada Y, Sugiyama H, Sugimoto A, Nojiri H, Sashihara K, Yokote K, Shimizu T (2018) Apple procyanidins promote

mitochondrial biogenesis and proteoglycan biosynthesis in chondrocytes. Sci Rep 8:7229. https://doi.org/10.1038/s41598-018-25348-1

McClure KA, Gong Y, Song J, Vinqvist-Tymchuk M, Palmer LC, Fan L, Burgher-MacLellan K, Zhang ZQ, Celton J-M, Forney CF, Migicovsky Z, Myles S (2019) Genome-wide association studies in apple reveal loci of large effect controlling apple polyphenols. Hortic Res 6:107. https://doi.org/10.1038/s41438-019-0190-y

McGhie TK, Walton MC (2007) The bioavailability and adsorption of anthocyanins: towards a better understanding. Mol Nutr Food Res 51:702–713. https://doi.org/10.1002/mnfr.200700092

Mellidou I, Chagné D, Laing W, Keulemans J, Davey MW (2012) Allelic variation in paralogues of GDP-L-galactose phosphorylase is a major determinant of vitamin C concentrations in apple fruit. Plant Physiol 160:1613–1629. https://doi.org/10.1104/pp.112.203786

Mendoza-Wilson AM, Castro-Arredondo SI, Espinosa-Plascencia A, del Refugio Robles-Burgueño M, Balandrán-Quintana RR, del Carmen Bermúdez-Almada M (2016) Chemical composition and antioxidant-prooxidant potential of a polyphenolic extract and a proanthocyanidin-rich fraction of apple skin. Heliyon 2:e00073

Mignard P, Beguería S, Reig G, Font i Forcada C, Moreno MA (2021) Genetic origin and climate determine fruit quality and antioxidant traits on apple (*Malus x domestica* Borkh.). Sci Hortic 285:110142. https://doi.org/10.1016/j.scienta.2021.110142

Min J, Li X, Huang K, Tang H, Ding X, Qi C, Qin X, Xu Z (2015) Phloretin induces apoptosis of non-small cell lung carcinoma A549 cells via JNK1/2 and p38 MAPK pathways. Oncol Rep 34: 2871–2879

Mosel HD, Herrmann K (1974) Changes in catechins and hydroxycinnamic acid derivatives during development of apples and pears. J Sci Food Agric 25:251–256

Na W, Ma B, Shi S, Chen Y, Zhang H, Zhan Y, An H (2020) Procyanidin B1, a novel and specific inhibitor of Kv10.1 channel, suppresses the evolution of hepatoma. Biochem Pharmacol 178: 114089

Nasir A, Bullo MMH, Ahmed Z, Imtiaz A, Yaqoob E, Jadoon M, Ahmed H, Afreen A, Yaqoob S (2020) Nutrigenomics: epigenetics and cancer prevention: a comprehensive review. Crit Rev Food Sci Nutr 60:1375–1387

Nezbedova L, McGhie T, Christensen M, Heyes J, Nasef NA, Mehta S (2021) Onco-preventive and chemo-protective effects of apple bioactive compounds. Nutrients 13(11):4025. https://doi.org/10.3390/nu13114025

Nguyen LT, Lee YH, Sharma AR, Park JB, Jagga S, Sharma G, Lee SS, Nam JS (2017) Quercetin induces apoptosis and cell cycle arrest in triple-negative breast cancer cells through modulation of Foxo3a activity. Korean J Physiol Pharmacol 21:205–213

Nile A, Nile SH, Shin J, Park G, Oh J-W (2021) Quercetin-3-glucoside extracted from apple pomace induces cell cycle arrest and apoptosis by increasing intracellular ROS levels. Int J Mol Sci 22:10749. https://doi.org/10.3390/ijms221910749

O'Rourke D (2021) Economic importance of the apple industry. In: Korban SS (ed) The apple genome. Springer, Cham, pp 1–18

Ou K, Gu L (2014) Absorption and metabolism of proanthocyanidins. J Funct Foods 7:43–53

Pereyra-Vergara F, Olivares-Corichi IM, Perez-Ruiz AG, Luna-Arias JP, García-Sánchez JR (2020) Apoptosis induced by (−)-epicatechin in human breast cancer cells is mediated by reactive oxygen species. Molecules 25:1020

Plunkett BJ, Henry-Kirk R, Friend A, Diack R, Helbig S, Mouhu K, Tomes S, Dare AP, Espley RV, Putterill J, Allan AC (2019) Apple B-box factors regulate light-responsive anthocyanin biosynthesis genes. Sci Rep 9:17762. https://doi.org/10.1038/s41598-019-54166-2

Podwyszyńska M, Marasek-Ciołakowska A (2021) Ploidy, genome size, and cytogenetics of apple. In: Korban SS (ed) The apple genome. Springer, Cham, pp 47–71

Rana S, Bhushan S (2016) Apple phenolics as nutraceuticals: assessment, analysis and application. J Food Sci Technol 53:1727–1738

Raudone L, Raudonis R, Liaudanskas M, Janulis V, Viskelis P (2017) Phenolic antioxidant profiles in the whole fruit, flesh and peel of apple cultivars grown in Lithuania. Sci Hortic 216:186–192

Renard C, Baron A, Guyot S, Drilleau J (2001) Interactions between apple cell walls and native apple polyphenols' quantification and some consequences. Int J Biol Macromol 29:115–125. https://doi.org/10.1016/S0141-8130(01)00155-6

Renard CM, Dupont N, Guillermin P (2007) Concentrations and characteristics of procyanidins and other phenolics in apples during fruit growth. Phytochemistry 68:1128–1138

Reyes-Farias M, Carrasco-Pozo C (2019) The anti-cancer effect of quercetin: molecular implications in cancer metabolism. Int J Mol Sci 20:3177

Romieu I, Varraso R, Avenel V, Leynaert B, Kauffmann F, Clavel-Chapelon F (2006) Fruit and vegetable intakes and asthma in the E3N study. Thorax 61:209–215

Roy S, Mondru AK, Chakraborty T, Das A, Dasgupta S (2022) Apple polyphenol phloretin complexed with ruthenium is capable of reprogramming the breast cancer microenvironment through modulation of PI3K/Akt/mTOR/VEGF pathways. Toxicol Appl Pharmacol 434: 115822. https://doi.org/10.1016/j.taap.2021.115822

Rupasinghe VHP, Huber GM, Embree C, Forsline PL (2010) Red-fleshed apple as a source for functional beverages. Can J Plant Sci 90:95–100

Sapio L, Salzillo A, Illiano M, Ragone A, Spina A, Chiosi E, Pacifico S, Catauro M, Naviglio S (2020) Chlorogenic acid activates ERK1/2 and inhibits proliferation of osteosarcoma cells. J Cell Physiol 235:3741–3752

Scafuri B, Marabotti A, Carbone V, Minasi P, Dotolo S, Facchiano A (2016) A theoretical study on predicted protein targets of apple polyphenols and possible mechanisms of chemoprevention in colorectal cancer. Sci Rep 6:32516. https://doi.org/10.1038/srep32516

Selby-Pham SN, Miller RB, Howell K, Dunshea F, Bennett LE (2017) Physicochemical properties of dietary phytochemicals can predict their passive absorption in the human small intestine. Sci Rep 7:1931

Selma MV, Espín JC, Tomas-Barberan F (2009) Interaction between phenolics and gut microbiota: role in human health. J Agric Food Chem 57:6485–6501

Shankar E, Zhang A, Franco D, Gupta S (2017) Betulinic acid-mediated apoptosis in human prostate cancer cells involves p53 and nuclear factor-kappa B (NF-κB) pathways. Molecules 22:264

Shilpi A, Parbin S, Sengupta D, Kar S, Deb M, Rath SK, Pradhan N, Rakshit M, Patra SK (2015) Mechanisms of DNA methyltransferase-inhibitor interactions: procyanidin B2 shows new promise for therapeutic intervention of cancer. Chem Biol Interact 233:122–138

Song W, Zhao X, Xu J, Zhang H (2017) Quercetin inhibits angiogenesis-mediated human retinoblastoma growth by targeting vascular endothelial growth factor receptor. Oncol Lett 14:3343–3348

Sotler R, Poljšak B, Dahmane R, Jukić T, Jukić DP, Rotim C, Trebše P, Starc A (2019) Prooxidant activities of antioxidants and their impact on health. Acta Clin Croat 58:726–736

Starowicz M, Achrem-Achremowicz B, Piskula M, Zielinski H (2020) Phenolic compounds from apples: reviewing their occurrence, absorption, bioavailability, processing, and antioxidant activity – a review. Polish J Food Nutr Sci 70:321–336

Stone WL, Pham T (2022) Biochemistry: antioxidants. STATPEARLS. https://www.statpearls.com/ArticleLibrary/viewarticle/17703

Stracke BA, Rüfer CE, Weibe FP, Bub A, Watzl B (2009) Three-year comparison of the polyphenol contents and antioxidant capacities in organically and conventionally produced apples (*Malus domestica* Borkh. cultivar 'Golden Delicious'). J Agric Food Chem 57:4598–4605

Stushnoff C, McSay AE, Forseline PL, Luby J (2003) Diversity of phenolic antioxidant content and radical scavenging capacity in the USDA apple germplasm core collection. Acta Hortic 623: 305–311

Sut S, Zengin G, Maggi F, Malagoli M, Dall'Acqua S (2019) Triterpene acid and phenolics from ancient apples of Friuli Venezia Giulia as nutraceutical ingredients: LC-MS study and in vitro activities. Molecules 24:1109. https://doi.org/10.3390/molecules24061109

Takos AM, Jaffé FW, Jacob SR, Bogs J, Robinson SP, Walker AR (2006) Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. Plant Physiol 142:1216–1232. https://doi.org/10.1104/pp.106.088104

Teh SL, Kostick SA, Evans KM (2021) Genetics and breeding of apple scions. In: Korban SS (ed) The apple genome. Springer, Cham, pp 73–104

Thilakarathna SH, Rupasinghe HPV, Needs PW (2013) Apple peel bioactive rich extracts effectively inhibit in vitro human LDL cholesterol oxidation. Food Chem 138:463–470. https://doi.org/10.1016/j.foodchem.2012.09.121

Thomas P, Dong J (2021)(−)-Epicatechin acts as a potent agonist of the membrane androgen receptor, ZIP9 (SLC39A9), to promote apoptosis of breast and prostate cancer cells. J Steroid Biochem Mol Biol 211:105906

Tian J, Zhang J, Han, Z-Y, Song T-T, Li J-Y, Wang Y-R, Yao Y-C (2017) McMYB12 transcription factors co-regulate proanthocyanidin and anthocyanin biosynthesis in *Malus* crabapple. Sci Rep 7:43715. https://doi.org/10.1038/srep43715

Tsao R, Yang R, Young JC, Zhu H (2003) Polyphenolic profiles in eight apple cultivars using high-performance liquid chromatography (HPLC). J Agric Food Chem 51:6347–6353

Ulaszewska M, Vázquez-Manjarrez N, Garcia-Aloy M, Llorach R, Mattivi F, Dragsted LO, Praticò G, Manach C (2018) Food intake biomarkers for apple, pear, and stone fruit. Genes Nutr 13:29. https://doi.org/10.1186/s12263-018-0620-8

van der Sluis A, Dekker M, de Jager A, Jongen W (2001) Activity and concentration of polyphenolic antioxidants in apple: effect of cultivar, harvest year, and storage conditions. J Agric Food Chem 49:3606–3613

van der Sluis A, Dekker M, Skrede G, Jongen W (2002) Activity and concentration of polyphenolic antioxidants in apple juice. 1. Effect of existing production methods. J Agric Food Chem 50: 7211–7219

van Nocker S, Berry G, Najdowski J, Roberto Michelutti R, Margie Luffman M, Philip Forsline P, Alsmairat N, Beaudry R, Nair MG, Ordidge M (2012) Genetic diversity of red-fleshed apples (*Malus*). Euphytica 185:281–293. https://doi.org/10.1007/s10681-011-0579-7

Veberic R, Trobec M, Herbinger K, Hofer M, Grill D, Stampar F (2005) Phenolic compounds in some apple (*Malus domestica* Borkh.) cultivars of organic and integrated production. J Sci Food Agric 85:1687–1694. https://doi.org/10.1002/jsfa.2113

Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, Salvi S, Pindo M, Baldi P, Castelletti S, Cavaiuolo M, Coppola G, Costa F, Cova V, Ri AD, Goremykin V, Komjanc M, Longhi S, Magnago P, Malacarne G, Malnoy M, Micheletti D, Moretto M, Perazzolli M, Si-Ammour A, Vezzulli S, Zini E, Eldredge G, Fitzgerald LM, Gutin N, Lanchbury G, Macalma T, Mitchell JT, Reid J, Wardell B, Kodira C, Chen Z, Desany B, Niazi F, Palmer M, Koepke T, Jiwan D, Schaeffer S, Krishnan V, Wu C, Chu VT, King ST, Vick J, Tao Q, Mraz A, Stormo A, Stormo K, Bogden R, Ederle D, Stella A, Vecchietti A, Kater MM, Masiero S, Lasserre P, Lespinasse Y, Allan AC, Bus V, Chagné D, Crowhurst RN, Gleave AP, Lavezzo E, Fawcett JA, Proost S, Rouzé P, Sterck L, Toppo S, Lazzari B, Hellens RP, Durel C-E, Gutin A, Bumgarner R, Gardiner SE, Skolnick M, Egholm M, Van de Peer Y, Salamini F, Viola R (2010) The genome of the domesticated apple (*Malus* × *domestica* Borkh.). Nat Genet 42:833–839

Verdu CF, Guyot S, Childebrand N, Bahut M, Celton J-M, Gaillard S et al (2014) QTL analysis and candidate gene mapping for the polyphenol content in cider apple. PLoS One 9(10):e107103. https://doi.org/10.1371/journal.pone.0107103

Vimolmangkang S, Han Y, Korban SS (2013) An apple MYB transcription factor, MdMYB3, is involved in regulation of anthocyanin biosynthesis and flower style/stigma development. BMC Plant Biol 13:176

Vimolmangkang S, Zheng D, Han Y, Khan MA, Soria-Guerra RE, Korban SS (2014) Transcriptome analysis of apple fruit excocarp identifies light-induced genes involved in red color pigmentation. Gene 534:78–87

Vinayagam R, Xu B (2015) Antidiabetic properties of dietary flavonoids: a cellular mechanism review. Nutr Metab 12:60

Volz RK, McGhie TK (2011) Genetic variability in apple fruit polyphenol composition in *Malus* x *domestica* and *Malus sieversii* germplasm grown in New Zealand. J Agric Food Chem 59: 11509–11521. https://doi.org/10.1021/jf202680h

Wang V, Chen X (2021) Genetics and genomics of fruit color development in apple. In: Korban SS (ed) The apple genome. Springer, Cham, pp 271–295

Wang X, Li C, Liang D, Zou Y, Li P, Ma F (2015) Phenolic compounds and antioxidant activity in red-fleshed apples. J Funct Foods 18(Part B):1086–1094

Wang N, Jiang S, Zhang Z, Fang H, Xu H, Wang Y, Chen X (2018) *Malus sieversii*: the origin, flavonoid synthesis mechanism, and breeding of red-skinned and red-fleshed apples. Hortic Res 5:70. https://doi.org/10.1038/s41438-018-0084-4

Wang L, Du H, Chen P (2020) Chlorogenic acid inhibits the proliferation of human lung cancer A549 cell lines by targeting annexin A2 in vitro and in vivo. Biomed Pharmacother 131:110673

Williams BA, Grant LJ, Gidley MJ, Mikkelsen D (2017) Gut fermentation of dietary fibres: physico-chemistry of plant cell walls and implications for health. Int J Mol Sci 18:2203

Wojdyło A, Oszmiański J (2020) Antioxidant activity modulated by polyphenol contents in apple and leaves during fruit development and ripening. Antioxidants 9:1–12

Wojdyło A, Oszmiański J, Laskowski P (2008) Polyphenolic compounds and antioxidant activity in new and old apple varieties. J Agric Food Chem 56:6520–6530

Wojdyło A, Nowicka P, Turkiewicz IP, Tkacz K, Hernandez F (2021) Comparison of bioactive compounds and health promoting properties of fruits and leaves of apple, pear and quince. Sci Rep 11:20253. https://doi.org/10.1038/s41598-021-99293-x

Wolfe K, Liu RH (2003) Apple peels as a value-added food ingredient. J Agric Food Chem 51: 1676–1683

Wood LG, Gibson PG (2009) Dietary factors lead to innate immune activation in asthma. Pharmacol Ther 123:37–53. https://doi.org/10.1016/j.pharmthera.2009.03.015

Wruss J, Lanzerstorfer P, Huemer S, Himmelsbach M, Mangge H, Höglinger O, Weghuber D, Weghuber J (2015) Differences in pharmacokinetics of apple polyphenols after standardized oral consumption of unprocessed apple juice. Nutr J 14:1–11

Wu K-H, Ho C-T, Chen Z-F, Chen L-C, Whang-Peng J, Lin T-N, Ho Y-S (2017) The apple polyphenol phloretin inhibits breast cancer cell migration and proliferation via inhibition of signals by type 2 glucose transporter. J Food Drug Anal 26:221–231

Xie X-B, Li S, Zhang R-F, Zhao J, Chen Y-C, Zhao Q, Yao Y-X, You C-X, Zhang X-S, Hao Y-J (2012) The bHLH transcription factor MdbHLH3 promotes anthocyanin accumulation and fruit colouration in response to low temperature in apples. Plant Cell Environ 35:1884–1897. https://doi.org/10.1111/j.1365-3040.2012.02523.x

Xu H, Wang N, Liu J, Qu C, Wang Y, Jiang S, Lu N, Wang D, Zhang Z, Chen X (2017) The molecular mechanism underlying anthocyanin metabolism in apple using the *MdMYB16* and *MdbHLH33* genes. Plant Mol Biol 94:149–165. https://doi.org/10.1007/s11103-017-0601-0

Xu M, Gu W, Shen Z, Wang F (2018a) Anticancer activity of phloretin against human gastric cancer cell lines involves apoptosis, cell cycle arrest, and inhibition of cell invasion and JNK signalling pathway. Med Sci Monit 24:6551. https://doi.org/10.12659/MSM.910542

Xu Y, Shu B, Tian Y, Wang G, Wang Y, Wang J, Dong Y (2018b) Oleanolic acid induces osteosarcoma cell apoptosis by inhibition of Notch signaling. Mol Carcinog 57:896–902

Yan Y, Liu N, Hou N, Dong L, Li J (2017) Chlorogenic acid inhibits hepatocellular carcinoma in vitro and in vivo. J Nutr Biochem 46:68–73

Yap WH, Lim YM (2015) Mechanistic perspectives of maslinic acid in targeting inflammation. Biochem Res Int 2015:279356

Zeng A, Liang X, Zhu S, Liu C, Wang S, Zhang Q, Zhao J, Song L (2021) Chlorogenic acid induces apoptosis, inhibits metastasis and improves antitumor immunity in breast cancer via the NF-κB signaling pathway. Oncol Rep 45:717–727

Zhang J, Xu H, Wang N, Jiang S, Fang H, Zhang Z, Yang G, Wang Y, Su M, Xu L, Chen X (2018) The ethylene response factor MdERF1B regulates anthocyanin and proanthocyanidin biosynthesis in apple. Plant Mol Biol 98:205–218

Zhang L, Hu J, Han X, Li J, Gao Y, Richards CM, Zhang C, Tian Y, Liu G, Gul H, Wang D, Tian Y, Yang C, Meng M, Yuan G, Kang G, Wu Y, Wang K, Zhang H, Wang D, Cong P (2019a) A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. Nat Commun 10:1494

Zhang L, Virgous C, Si H (2019b) Synergistic anti-inflammatory effects and mechanisms of combined phytochemicals. J Nutr Biochem 69:19–30

Zhang B, Yang HJ, Yang YZ, Zhu ZZ, Li YN, Qu D, Zhao ZY (2020) mdm-miR828 participates in the feedback loop to regulate anthocyanin accumulation in apple peel. Front Plant Sci 11: 608109. https://doi.org/10.3389/fpls.2020.608109

Zhang B, Yang H-J, Qu D, Zhu Z-Z, Yang Y-Z, Zhao Z-Y (2022) The MdBBX22–miR858–*MdMYB9/11/12* module regulates proanthocyanidin biosynthesis in apple peel. Plant Biotech J 20:1683–1700

Zhao C-N, Meng X, Li Y, Li S, Liu Q, Tang G-Y, Li H-B (2017) Fruits for prevention and treatment of cardiovascular diseases. Nutrients 9:598

Zhou K, Hu L, Li P, Gong X, Ma F (2017) Genome-wide identification of glycosyltransferases converting phloretin to phloridzin in *Malus* species. Plant Sci 265:131–145

Zhu Y, Evans KM, Peace C (2011) Utility testing of an apple skin color *MdMYB1* marker in two progenies. Mol Breed 27:525–532

Zhu W, Li MC, Wang FR, Mackenzie GG, Oteiza PI (2020) The inhibitory effect of ECG and EGCG dimeric procyanidins on colorectal cancer cells growth is associated with their actions at lipid rafts and the inhibition of the epidermal growth factor receptor signaling. Biochem Pharmacol 175:113923

Zielinska D, Laparra-Llopis JM, Zielinski H, Szawara-Nowak D, Giménez-Bastida JA (2019) Role of apple phytochemicals, phloretin and phloridzin, in modulating processes related to intestinal inflammation. Nutrients 11:1173

Zou T, Wang B, Li S, Liu Y, You J (2020) Dietary apple polyphenols promote fat browning in high-fat diet-induced obese mice through activation of adenosine monophosphate-activated protein kinase α. J Sci Food Agric 100:2389–2398

Integrating Omic Tools to Design Nutraceutically Rich Citrus

Bidisha Mondal

Contents

Abstract

The world is facing a bipartite problem of malnutrition covering under-nutrition as well as over-nutrition. This situation is an outcome of imbalance in financial capability and inequality in food distribution. The intake of diverse nutraceutical compounds could assist in overcoming the health problems and may provide additional protection to the human communities through generation of substantial antimicrobial and antioxidant properties. The tropical and subtropical fruits and vegetables are rich source of multiple bioactive compounds and are regarded as therapeutic storehouse. In majority of the countries, the fruit by-products and

B. Mondal (✉)
Department of Genetics and Plant Breeding, The School of Agriculture and Allied Sciences, The Neotia University, Sarisha, West Bengal, India
e-mail: bidisha.mondal@tnu.in

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_35

wastes lead to a complication of agro-waste disposal and management. The fruits, as a whole, are fascinating in production of bioactive elements. Citrus is a representative fruit having marked significance in food and fragrance sector. The nonedible parts, including peel, seed, fiber, and leaves, are potent source of health-promoting elements. The sequencing of the Citrus genome and available omics information could distinctly aid in the development of a nutraceutical database. In this chapter, the contemporary account on Citrus omics, encompassing genomics, transcriptomics, and metabolomics are critically evaluated for development of an array of sub-foods pertinent to food-omics division. The Citrus germplasm recourse and exploitable breeding strategies available till date could build a designer Neutri-Citrus appropriate for the prospering nutraceutical sector. This futuristic strategy of utilization of Citrus fruit waste and by-products in phyto-pharmaceutical industry may simultaneously encourage global agro-waste reduction and emerge as a new model contributing to circular economy.

Keywords

Citrus · Omics-tools · Nutraceutical · Germplasm · Designer food · Agro-waste · Circular economy

1 Introduction

Nutraceuticals are natural dietary components found in food having health benefit and disease-fighting property. Nutraceuticals exhibit a boundary between food and drug and could be exploited to ameliorate health, slow down senescence, and maintain proper functioning of human body. A broad functional classification of nutraceuticals includes two types: potential nutraceutical and established nutraceutical. A prospective sub-food could be considered as a potential nutraceutical and could be elevated to an established designer food subject to qualifying certain clinical trials. In current time, GRAS (generally recognized as safe) status of nutraceuticals is essential for the quality assurance of products. The common food sources used as nutraceuticals are covered under multiple categories. The dietary fiber, probiotics, and prebiotics contain elements for improvement of gut health. Polyunsaturated fatty acids (PUFAs), polyphenols, antioxidants, vitamins, and spices form a category with capacity of fighting free radical led damages (Das et al. 2012).

Plants harbor bioactive compounds with health benefits beyond the basic nutritional values in the form of secondary metabolites. On the basis of chemical structure and function, the common secondary metabolites available in foods could be grouped as phytosterols, phytoestrogen, carnitine, and choline with direct cellular and tissue-specific activities. Carotenoids, dithiolthiones, flavonoids, polyphenols, glucosinolates, and taurine are equally potent modulators of physiological processes (Rodríguez-Casado 2014). Fruits and vegetables with diverse colors are quintessential source of an array of nutraceuticals. The classical examples are grape (*Vitis vinifera*), watermelon

(*Citrullus lanatus*), and banana (*Musa* spp.). Fruits like bael (*Aegle marmelos*), pomegranate (*Punica granatum*), amla-Indian gooseberry (*Phyllanthus emblica*), cranberry (*Vaccinium* spp.), orange (*Citrus sinensis*), and lemon (*Citrus limon*) are well-established sources of nutraceuticals (Dutta et al. 2017). Citrus fruits including orange, bergamots (*Citrus bergamia*), lemon, and grapefruit (*Citrus x paradisi*) are rich in bioactive compounds. The tropical and subtropical fruits are considered as therapeutic storehouse. The pharmaceutical and nutraceutical components present in fruits protect human from several diseases, increases life expectancy and reduces stress. The fruit peel, seeds, and by-products from food processing units serve as potential source of multiple bioactive compounds. These waste products could be utilized for extraction and valorization of nutraceuticals that strengthen the field of circular economy and aid in environmental waste mitigation (Fig. 1).

The phytochemistry of Citrus fruits and their by-products has gained importance in nutraceutical industry due to its impact on human health and contribution toward nutragenomics. The human gene expression, transcriptomic modulation, and protein turn-over could be regulated by Citrus bioactive compounds. In specific instances, the expression of *COX2*, microsomal cytochrome *P450 A1*, and *NF-kB* genes were guided by hesperidin, naringenin, and hesperetin, nutraceuticals that are present in Citrus. Evidence reveals that Citrus flavonoids have a regulatory effect on gene expression by coding of low-density lipoprotein receptors (LDLR). Citrus essential oil is also well known for food preservation with substantial antimicrobial, flavoring, and antioxidant properties (Mahato et al. 2018). The Citrus peel powder possesses free radical scavenging activity and could replace synthetic preservatives in the market in near future (Khan et al. 2021).

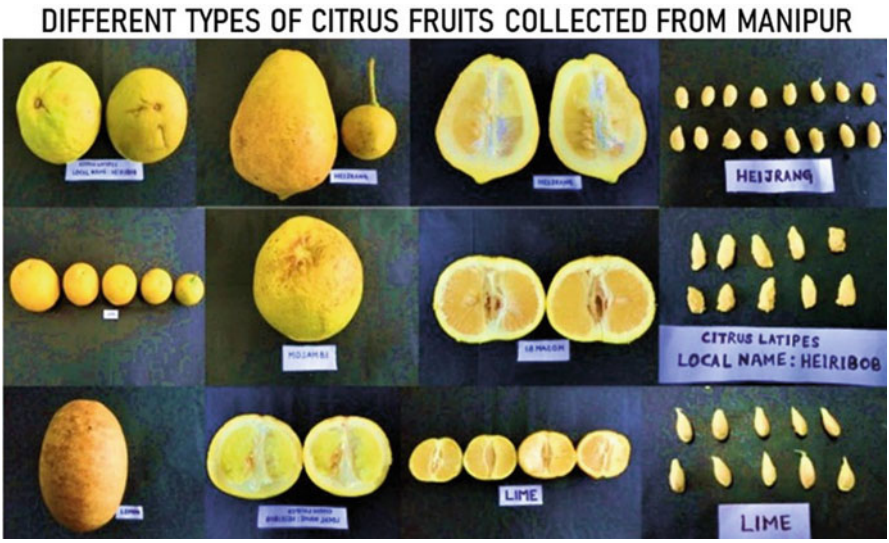


Fig. 1 Citrus fruits collected from Manipur, India (*C. latipes*, *C. medica*, *C aurantifolia*)

2 Mechanism of Nutraceutical Production in Citrus

The qualitative and quantitative variation in bioactive compounds produced in vegetatively grown plants bears epigenetic regulation. The plant genome controls production of the expected phyto-constituents, but epigenomic events change the ultimate phenotypic imagery of the genotype. Parental imprinting and epi-alleles could affect developmental switches leading to gene silencing or over-expression in some natural ecotypes, producing variants. In clonally propagated plants, the recovery of somaclones is possible due to reprogramming of the somatic embryos, carrying similar genetic architecture as sib-plants. The movement of diffusible epigenetic signals through plasmodesmata and vascular systems as well as parasites from one part to another conceivably regulates the fitness of a plant. The season, ecology, and light intensity might regulate clonally propagated plant to a larger extent than its sexually grown complements (Pikaard and Scheid 2014). Several exogenous as well as endogenous stimuli determine the survival and nutraceutical production of various genotypes (Fig. 2).

The by-products of the Citrus industry are of immense economic value. The red orange of Sicily (known as blood orange) is a reservoir of significant amount of biologically active compounds (limonoids and flavonoids). The decanted pulps were the most abundant source of highest amount of flavonoids (130 g/kg) and a high amount of limonoids (5.5 g/kg) earning the protected geographic indication (PGI) for the cultivar. In Citrus, seeds are the best source of limonoids with about 10 g/kg of expression. Low amount of anthocyanins were found only in coarse pulps and waste water of red oranges (Russo et al. 2021). Bergamot (*Citrus bergamia* Risso) contains limonoids in abundance in seeds and peels (70% and 80% of the total, respectively), while limonoid glucosides are more abundant in juices and pulps

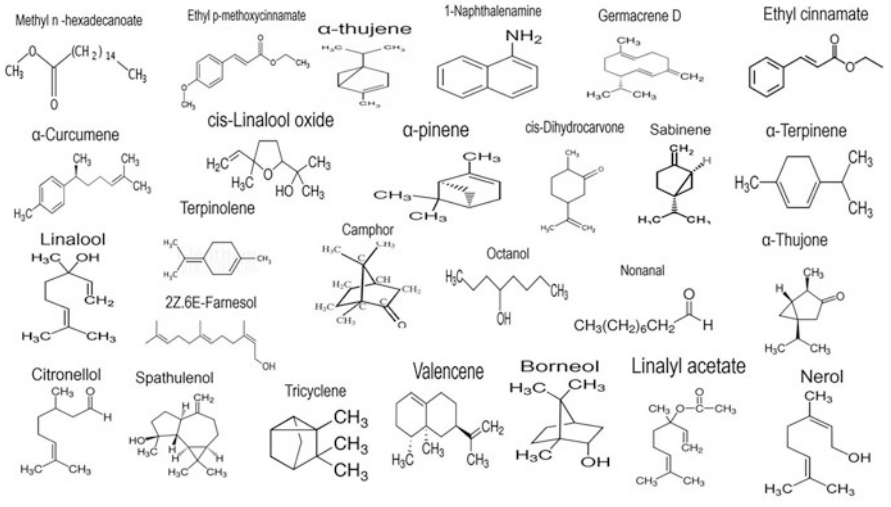


Fig. 2 Major bioactive components present in Citrus peel essential oil

(61% and 76% of the total, respectively). The limonoids are related to microbial disintegration and free radical scavenging activity (Russo et al. 2016). *C. sinensis* commonly called 'Ovale Calabrese' grows profusely in the south of Italy (Calabria region). Due to the richness of flavonoids and limonoids and high content of phenolic compounds, in particular hesperidin in the pastazzo, the fruits could be utilized for the valorization of the by-product for the development of a novel functional ingredient (Celano et al. 2019).

The lemon leaf essential oil (EO) is rich in oxygenated monoterpenes, accounting over 48% of the extracted oil. The most abundant constituent of the oil is a terpenoid identified as neryl acetate (15.3%). Geranyl acetate is a monoterpenoid and its alcohol nerol and geraniol are involved in the generation of the characteristic lemon aroma. The sesquiterpene hydrocarbons fraction in the EO was represented by β-caryophyllene representing the second most quantitatively important chemical class of compounds. Traces of β-bisabolene and bicyclogermacrene were detected in lemon EO. Principle monoterpene hydrocarbons limonene and β-pinene along with spathulenol and caryophyllene oxide, the oxygenated sesquiterpenes were found to be a possible contributor in the lemon EO. The essential oil composition of lemon leaves exhibits great variability in onto-genetics. Leaves of lemon specimens from the Mediterranean region were mainly rich in limonene with the aldehydes and acetic esters of both nerol and geraniol (Vekiari et al. 2002), while the composition of Indian lemon leaf EO exhibits the presence of (Z)-sabinene hydrate, geraniol, with a trace of α-pinene (Pal et al. 2016). The Iranian lemon oil is predominated by linalool, followed by a substantial representation of geraniol, α-terpineol, as well as linalyl acetate (Hojjati and Barzegar 2017). The geographical origin of the genotypes plays an instrumental role in the composition of the essential oil (EO) extracted from the Citrus peel.

The orange leaf essential oil was mainly dominated by oxygenated monoterpenes. The most prevalent component for this class was linalool. The acetic ester of linalool, α-terpineol, geranyl, and neryl acetate was associated with the signature aroma of the fruit. Myrcene, β-ocimene, and β-pinene belonging to monoterpene hydrocarbons were another related chemical class of volatiles found in orange fruits. The metabolic profile of oranges expresses differential outcomes subject to growth conditions and nurturing. The essential oil profile of different Citrus accessions under orange is consistent with the literature report presented earlier (Sanmartin et al. 2019).

The leaf essential oil of endemic species of Australia reveals significant variation in oil content. While *C. australasica* produced oil is predominated by bicyclogermacrene, germacrene-D, δ-elemene, and limonene, the other promising species, *C. australis* oil shows major share of α-pinene. The oil from another species, *C. garrawayi*, shows two distinct components, α-pinene and β-caryophyllene, belonging to monoterpene group. *C. glauca* contains α and β-pinene in contrast to *C gracilis* with γ-terpinene as major contributor. Another species, *C. inogora*, contains germacrene D as principal constituent. The wide variation of component availability is a comprehensible reflection of disparate epigenetic control of the same pathway (Brophy et al. 2001).

C. aurantifolia leaf and peel essential oil was characterized by 49 constituents (93.6% of the total oil), in which the dominant components were the monoterpene hydrocarbon limonene, β-pinene, γ-terpinene, and β-myrcene. *C. aurantium*

essential oil displays 15 components, with limonene as the major monoterpene hydrocarbon with successors β-myrcene and α-pinene. The oil has minor fractions of linalool and linalyl acetate, representing the oxygenated monoterpenes (Tundis et al. 2012). The oil profile of both the above-mentioned species could be utilized as natural anti-oxidants and could prevent an array of neurodegenerative diseases and aging. Bioactive compounds naturally occur in our surrounding plants, but in trace quantities. Epidemiological studies indicate that intake of bioactive food increases gastrointestinal ecology, immunity, reduces risk of cancer, diabetes, heart diseases, Alzheimer's disease, stroke, cataract, cytotoxicity, and even age-induced senescence (Saini et al. 2022). Additionally, energy boosting, wound recovery, and beauty retention are also ensured by nutraceuticals (Table 1). Limonoids exhibit antimicrobial and anticancer activities and is a recognized human health promoter with manifold pharmacological properties and potential (Brito et al. 2014).

3 Chemistry of Major Bioactive Compounds Present in Citrus

3.1 Flavonoids

Flavonoids are found in all parts of plants and form a class of polyphenolic secondary metabolites. Flavonoids contain a 15-carbon framework (C6-C3-C6). Two hexa-carbon phenyl rings form a heterocyclic ring with embedded oxygen. A variety of modifications in the heterocyclic ring of flavonoid generates several subgroups. The naringin, eriocitrin, hesperidin, and narirutin constitute flavanone group, while flavones were categorized into rhoifolin, vitexin, and diosmin. Polymethoxylated flavones with natural antioxidant property cover nobiletin, tangeritin, and 5-demethyl nobiletin. Flavonols, the colorless secondary metabolites housing kaempferol, quercetin, and rutin, are recognized developmental regulators and signaling receptors. The colorful pigment anthocyanin present inside vacuolar compartments provides numerous health benefits to humans. The anthocyanin covers the cyanidin and peonidin glucosides (Testai and Calderone 2017; Saini et al. 2022). Plant defense, cell signaling, and repair from UV associated damages were mainly regulated by the flavonoid group of bioactives.

3.2 Carotenoids and Apocarotenoids

Carotenoids are a universal accessory isoprenoid pigments involved in carbon fixation and signaling pathways. Pigment carotenoid produces two subgroups: carotenes and xanthophylls. The hydrocarbon carotenoids α-carotene, β-carotene, and lycopene act as immune-activator. While xanthophylls, the oxygenated derivatives of hydrocarbon carotenoids represented by neoxanthin, violaxanthin, lutein, and β-cryptoxanthin, were accepted as anti-phototoxic agents (Saini and Keum 2018). Apocarotenoid is

Table 1 Health benefit related to consumption of diverse Citrus fruit species

S. No	Plant part	Species	Medical application	Reference
1.	Peels	*Citrus maxima*	Diabetes, hypertension	Oboh and Ademosun (2011)
2.	Rind	*Citrus sinensis*	Lung cancer	Xiao et al. (2009)
3.	Sweet lime	*Citrus limetta*	Wound healing	Harsha und Aarti (2015)
4.	Peel	*Citrus paradisi*	Anti-inflammatory	Eneke et al. (2021)
5.	Fruit extract	*Citrus aurantifolia*	Diabetes mellitus	Silalahi (2002)
6.	Peel, pulp	*Citrus latifolia*	Antibacterial activity	Medina-Torres et al. (2019)
7.	Pulp	*Citrus grandis*	Colorectal, breast cancer	Tocmo et al. (2020)
8.	Peel oil	*Citrus jambhiri*	Anti-ulcer, antinociceptive	Babarinde et al. (2021)
9.	Peel	*Citrus unshui*	Edema, capillary leakage	Tsitsagi et al. (2018)
10.	Peel, pulp	*Citrus medica*	Anticatarrhal, analgesic	Chhikara et al. (2018)
11.	Peel	*Citrus aurantium*	Cellulite reduction	Jabri and Marzouk (2013)
12.	Plant extract	*Citrus limon*	Pneumonia, skin disorder	Shaikh et al. (2022)
13.	Fruit	*Citrus latipes*	Antioxidant, UV absorber	Rao et al. (2021)
14.	Rind extract	*Citrus hystrix*	Alzheimer's disease	Siti et al. (2022)
15.	Leaf oil	*Citrus limonimedica*	Anticancer	Lota et al. (1999)
16.	Fruit	*Citrus clementina*	Antidiabetic, antioxidant	Loizzo et al. (2018)
17.	Leaf oil	*Citrus reshni*	Anti-inflammatory	Hamdan et al. (2013)
18.	Fruit oil	*Citrus bergamia*	Urinary tract infection	Navarra et al. (2015)
19.	Leaf oil	*Citrus australasica*	Anticancer, foot skin disorder	Wang et al. (2019)
20.	Leaf oil	*Citrus glauca*	Antibacterial, antileukemic	Scora and Ahmed (1995)
21.	Rind oil	*Citrus ichangensis*	Anticholesterol, anticancer	Herman et al. (1989)
22.	Seed	*Citrus junos*	Antitumor	Shon and Park (2006)

another category of pigment detected in Citrus with distinctive role in signaling and growth regulation. The origin of apocarotenoids is subject to cleavage of dioxygenase. The eco-nutritional variation of apocarotenoid is an important factor for its acceptance as a functional food. In Citrus, both β-cryptoxanthin and zeaxanthin could be the probable progenitor of β-citraurin, and the process witnesses an asymmetric cleavage of the progenitor (Luan et al. 2020).

3.3 Terpenes and Limonoids

The monoterpene hydrocarbons consist of two isoprene components and cover several subtypes. The oxygenated monoterpenes (nonanal, geranial, and neral) are known as fragrant. Terpene alcohols are olfactory markers segregated into linalool, verbenol, geraniol, carveol, and α-terpineol. Sesquiterpenes with 15 carbon and 3 isoprene units are the major chemical constituents of the volatile fractions of Citrus essential oil (Raspo et al. 2020). The total terpenoids expressed in mandarin, tangerine, grapefruit, orange, citron, and lemon essential oil were preferentially dominated by a monoterpene D-limonene. An interesting study reveals that green mandarin contains fourfold higher limonin, a tetracyclic triterpenoidin much higher amount than its yellow and red complements. Montenegrin mandarin, essential oil has γ-terpinene as a major fraction with the minor presence of citronellol, an acyclic monoterpenoid, and terpene alcohol (Rossi et al. 2020). This sparse presence of these trace signatory components favored the antioxidant activity with a property of significant decrease of cytotoxicity inside colorectal cancer HT-29 cells.

3.4 Phenolic Acids

In a study involving kinnow mandarin clearly stated the percentage of solvent in determination of final recovery of the peel extracts or peel essential oil. Among the phenolic compounds, ferulic acid and hydroxycinnamic acid were recovered with remarkable antiaging properties. Hesperidin, a flavone glycoside, was also abundant in kinnow mandarin peel extracts. A tri-hydroxybenzoic acid, gallic acid, and catechin, a poly-phenol compounds were found to remain present in high concentration in kinnow type mandarin. Trace quantities of caffeic acid, a methylxanthine, and naringenin, a flavanone, were recovered from the peel extract with anti-carcinogenic and immune-stimulant activity (Safdar et al. 2017).

3.5 Coumarin

Five coumarins and 21 fucocoumarin compounds are widely present in Citrus species. The compounds are rich in dimethylallylated and/or geranylated compounds such as bergamottin, aurapten, or imperatorin. The coumarin compounds exhibit multiple applications in skin diseases with antiviral and anticancer properties. Citrus roots are storehouse of some novel coumarins. Peroxytamarin, with antibacterial property along with cis-casegravol having antiproliferative properties were detected as major constituents. Lignan glycoside, citrusarin-A, and citrusarin-B were present in root of Citrus plants. These polycyclic compounds possess organoleptic properties with prominence as satisfactory odor and flavor agents (Ito et al. 1991).

4 Citrus Genome and Phylogeny

Citrus is a diploid genus with 18 chromosomes and a genome size ranging from 265 to 407 Mb. In *Citrus* spp., a spectrum of research has been done on germplasm characterization and management. The Sino-Indian border area is regarded as a unique germplasm conservation center of Citrus. Citrus group contains multiple species qualifying as scions (commercial varieties) and several rootstocks (wild types). The novel *Citrus* fruit types of South Asia displays diverse species, including *C. reticulata*, *C. sinensis*, *C. aurantifolia*, *Citrus indica*, *Citrus limon*, *C. aurantium*, *C. maxima*, *C. medica*, *C. limetta*, *C. latipes*, *C. jambhiri*, and *C. hystrix* (Mondal 2021). These huge ecotypic resources could act as reservoir for a number of important bioactive compounds that could play a crucial role in Indian nutraceuticals industry. The North Eastern hilly region evinces *C. indica*, *C. assamensis*, *C. macroptera*, *C. latipes*, *C. ichangensis*, *C. micrantha*, *C. medica*, *Fortunella margarita*, *F. crassifolia*, *F. japonica*, and *Poncirus trifoliata.* The analysis of the pulp, peel, and seed of this gigantic germplasm reservoir, predominantly the wild types, could manufacture unique compounds with prospects in the medicinal and pharmaceutical sectors.

In Chinese wild mandarin Mangshanju (*Citrus reticulata* Blanco), a total of 81 compounds were identified, including flavonoid glycosides, acylated flavonoid glycosides, flavones, polymethoxylated flavonoids, and limonoids, as well as four other compounds. The Citrus wild germplasm reserves are the automatic store house of nutraceuticals, and a thorough characterization of the constituents could yield an optimum product for the industry. The wild plant produces 22 polymethoxylated flavones and 10 polymethoxylated flavanones/chalcones in excess in comparison to its commercially propagated counterpart (Zhao et al. 2018). Ancient Indian Citrus species, pummelo (*Citrus grandis* L. Osbeck) harbors naringin as the predominant flavonoids irrespective of genotypes. Some selections are particularly high in naringin, and others seem promising for lycopene and phenols. Other phenolics quantified in the juice included caffeic, epicatechin, benzoic acid, neoeriocitrin, hesperidin, and narirutin. The examples cited above present the importance of germplasm characterization. The proper characterization of the wide and rich gene-pool with a nutraceutical inventory of elite cultivars could possibly assist Citrus nutraceuticals breeding program (Nishad et al. 2018).

5 Omics Understanding of Nutraceutical Production in Citrus

The omics tapestry of living organisms presents the most mysterious and interesting field of molecular research. The response of a plant genome to external stresses and the successive cellular homoeostasis are regulated by the genic and inter-genic territories of the genome. In metabolic discovery, the natural variation within the species provides significant insight into nutraceuticals production. The multiple genomic identities of Citrus, including the wild and cultivated types, influence a

range of traits covering plant defense and commercially important attributes. Due to simultaneous occurrence of natural hybridization and vegetative propagation in citrus, the mode of inheritance of metabolite production becomes complex. Most of the biosynthetic pathways leading to identification of numerous bioactive compounds mainly follow non-Mendelian inheritance. The mQTL mapping technique unravels novel genes with intense epigenetic and environmental influence in synthesis of diverse end-products. The analysis and integration of various omics tools will assist in achieving futuristic goals in Citrus nutra-omics.

5.1 Nutragenomics of Citrus

In Citrus, the multidrug and toxic compound extrusion (*MATE*) gene plays a valuable role in flavonoids and citrate metabolism in fruits and is involved in numerous allied physiological processes. The MATE proteins are a class of secondary active multidrug transporters. In *Citrus clementina*, a total of 69 MATE transporters were classified on the basis of phylogeny. Tandem and segmental duplication events were the main causes of the Citrus MATE family expansion. RNA sequencing and qRT-PCR operations in different fruit developmental stages revealed that *CitMATE* gene expression is regulated by tissue diversity and a specific developmental stage. *CitMATE43* and *CitMATE66* are involved in the transport process of flavonoids and citrate in Citrus fruit, and additionally *CitERF32* and *CitERF33* are related to activation of *CitMATE43* promoter (Liu et al. 2022a).

In a research involving the large family of plant polyphenolic secondary metabolites, it has been expressed that transgenic plants could be developed with a high production capacity for bioactive compounds. A model transgenic tomato developed with structural flavonoid genes (encoding stilbene synthase, chalcone synthase, chalcone reductase, chalcone isomerase, and flavone synthase) from different plant sources was able to produce novel nutraceuticals. Biochemical analysis showed that the tomato peel contained high levels of stilbenes (resveratrol and piceid), deoxychalcones (butein and isoliquiritigenin), flavones (luteolin-7-glucoside and luteolin aglycon), and flavonols (quercetin glycosides and kaempferol glycosides). The study involving genetic engineering of flavonoids in tomato fruit demonstrated the prospects of developing tailor-made Citrus crops with high level of plant nutrients (Abeynayake et al. 2012).

In a research involving the identification of major limonoid-producing gene in Citrus, P450s (CYP450s), MYB, and CiOSC genes were found to play a ubiquitous role in limonoid biosynthesis. The leaves, phloem, and seeds of pummelo (*Citrus grandis* L. Osbeck) express variations in limonoids' contents at different development stages. Digital gene expression profiling identified 382 putative genes from 3 conjunctive groups related to the biosynthesis of limonoids. Repetitive correlation analysis with the samples from different genotypes involving different developing tissues of the Citrus confirmed 15 candidate genes with high correlation with the contents of limonoids. The cytochrome P450s (*CYP450s*) and transcriptional factor *MYB* demonstrated significantly high correlation coefficients, validating the

importance of those genes for the biosynthesis of limonoids. *CiOSC* gene encoding the critical enzyme oxidosqualene cyclase (OSC) for biosynthesis of the precursor of tri-terpene scaffolds was responsible for limonoids production in seeds. Suppressing the expression of *CiOSC* with *VIGS* (Virus-induced gene silencing) demonstrated that the level of gene silencing was significantly correlated to the reduction of limonoids contents (Wang et al. 2017).

In Citrus cultivars, two vacuolar P-ATPase homologs, *CitPH1* and *CitPH5*, are strongly induced in highly acidic species such as lime and lemon fruits. The level of expression of these two genes is significantly reduced in acidless mutants. The cellular pH gradient is maintained by vacuolar ATPases with an array of genes: *CitPH1*, *CitPH5*, *PH3* (WRKY), *PH4* (*MYB*), and basic helix-loop-helix (*bHLH*). Citrus Noemi (*CitAN1*) transcription factors were less expressed or reduced in the mutants (Butelli et al. 2019). Homologs of *PH4*, which is a *MYB* transcription factor, have been reported to activate the promoter of the proton pumps (*PH1* and *PH5*) in Citrus. A pleiotropic gene Noemi (*CitAN1*) regulates both acidless phenotypes and loss of proanthocyanidin and anthocyanin in sweet lime, citron, sweet orange, lemon, and limetta accessions. A variation in the core promoter region of the *Noemi* gene in two limetta accessions reduces gene expression and increases the pH of the juice, indicating *Noemi* as a vital gene contributing to fruit acidity. Recent studies showed that a specific mutation in *CitAN1* is responsible for the reduced expression of *CitPH1* and *CitPH5* in acidless lemon and other acidless Citrus fruits (Strazzer et al. 2019). A complex formed of WD40-repeat proteins, MYB, bHLH, and WRKY transcription factors known as WMBW in short plays a significant role in the operation of three anthocyanin pathways. *Ruby2* and *Ruby1* genes belonging to a cluster function as anthocyanin activators, but Ruby2 operates in tender leaves and Ruby1 functions preferentially in fruits. The reduction or loss of anthocyanins may be due to the hitchhiking effect of fruit acidity selection in some domesticated types. A mutant gene, *AN1*, is a common regulator for both citric acid and anthocyanin metabolism, playing a regulatory role in fruit acidity and pigment production (Rao et al. 2021).

The National Centre for Biotechnology Information (NCBI) reveals enormous data on Citrus, categorized under 20 databases. The Citrus literature is covered under 42 bookshelves with 6 NML catalog, 2407 Pubmed, 17,780 Pubmed central information (Table 2). The data available in NCBI exhibits 110,766 terpinene, 382 limonoid, 259 alcohol dehydrogenase, 162 flavonoid, 32 limonene, 22 γ-terpinene, 67 coumarin, 60 chalcone synthase, 48 geraniol, 28 carotenoid, 8 citronellol, 7 citral, 7 quercetin, 6 trans-citral, 5 cis-citral, 5 naringenin, 3 anthocyanin gene, and 2 kaempferol associated genes in *Citrus* genera (NCBI 1988). These potential candidate genes could be utilized for qualitative as well as quantitative improvement of bioactive accumulation in pulp, peel, and seeds of Citrus.

5.2 Nutra-transcriptomics of Citrus

The accurate comprehension of the biosynthetic pathway is essential for commercial metabolite production. Transcriptomic data-mining is an efficient tool for

Table 2 Citrusomics and related databases present in NCBI platform

Citrus (taxonomy ID: 2706)	
Name of the database	Reported information
Domain: genome	
Genome	28
Assembly	76
Taxonomy	1
Nucleotide	1,230,115
Bio-collection	3
SRA	13,928
Domain: gene	
Gene	110,766
Geo-data set	2316
Pop-set	2365
Domain: protein	
Protein	1,299,325
Protein family model	13
Conserved domains	11
Identical protein groups	352,797
Structure	59
Domain: pathways	
Pathways	1320
Substances	360
Bioassays	640
Compounds	14
Domain: clinical	
Clinical trials.gov	191
dbGaP	5
MedGen	2
OMIM	1

identification of gene families involved in certain metabolite production. In non-model plant species, a transcriptomic approach could accelerate genomic investigation. Transcriptomic analysis could efficiently identify the genetic and epigenetic factors underlying the expression of diverse bioactive compounds present in various plant species and the interplay of gene and environment influencing distinct pathways. A study involving the transcriptomic profiling of two varieties of *Citrus reticulata* expressed differences in the storage of bioactive compounds. Transcriptomic profiling of *Citrus reticulata* 'Huajuhong' (HJH) and *C. reticulata* 'Sanhuhongju' (SHHJ) reveals differences in the distinct level of bioactive ingredients in fruit peels. The total flavonoid in HJH peels was significantly higher than that in SHHJ. Messenger RNA (m-RNA) sequencing identified 203 differentially accumulated metabolites (DAMs) and 3517 differentially expressed genes (DEGs). Among the DAMs, the major components were flavonoids (104, 51.2%), followed by phenolic acids (30, 14.7%). KEGG enrichment analysis indicated the over-

expression of the flavonoid pathway. Ten glucosyl transferase genes regulated the accumulation of seven of the top ten flavonoid glycosides in HJH (Yu et al. 2022). Collectively, the higher content of flavonoid glycosides in HJH peels than SHHJ might contribute to the distinct health-promoting effect of the former and confirms its possible selection for integration in nutraceutical extraction.

In a similar transcriptomics study conducted by a group of scientists involving blood orange (*Citrus sinensis* cv. 'Tarocco') showed significant effects of diverse exogenous treatments in regulation of internal qualities of the fruit. Bagging has been widely used in fruit crops to improve fruit quality, but the result showed that bagging treatment has a significant effect on fruit quality. The treatment led to increase in total flavonoid (TFL) and total anthocyanin (TAN) concentrations, while total soluble solids (TSS), total phenolics (TPH), ascorbic acid (AsA) concentrations, and titratable acidity (TA) decreased in response to the bagging treatment. The high-throughput tag-sequencing (Tag-seq) analysis detected over 21 × 106 clean reads per library. Approximately 53.7–71.7% genic and 3.1–6.4% of intergenic clean reads were mapped onto Citrus genomic regions, respectively. About 25.2–39.9% of the clean reads failed to align with the Citrus genome. Overall, bagging treatment resulted in an increase in transcripts involved in a range of metabolic pathways rather than anabolic pathways. The tricarboxylic acid (TCA) cycle, sucrose and starch metabolism, ascorbate metabolism, and the phenylpropanoid pathway got affected through the treatment. Competition for limited amounts of substrates for the flavonoid, phenolics, and anthocyanin pathways may have led to an increase in total anthocyanins (TAN) and total flavonoid (TFL) concentrations under a variety of stresses on fruits. The gene, *bHLH* is involved in anthocyanin biosynthesis in blood orange fruit and was identified as a key player in controlling the genetic mechanisms operating in Citrus fruit in response to a bagging treatment (Sun et al. 2014).

In another, transcriptomic study on cold-induced response of *Citrus paradisie* confirms that low temperature exposure induces the elevation in the anthocyanin levels in the orange flesh. This enhancement being modulated by the transcriptional stimulation of the genes involved in the anthocyanin biosynthesis. The RNA profiling and subsequent construction of expressed sequence tag (EST) collection revealed *NAC* family of gene plays a vital role in the abiotic stress response. An EST, encoding the BRD4 bromodomain, a DNA-binding protein belonging to an extensive family of evolutionarily conserved protein originally remain associated with chromatin activity. This BRD4 plays a pivotal role in chromatin remodeling and transcriptional activation. Similarly, another gene, glutathione transferase (*Tau2*), previously identified in the leaves of blood oranges, ensures ROS scavenging activity. The storage temperature as well as the exposure tenure turned out to be critical for pigment development in flesh tissue. Though prolonged cold storage for more than 3 months, negatively influences the sensory quality of oranges due to the increase of the malodorous substance vinylphenol, whereas the standard fruit quality parameters (total soluble solids (TSS), total acidity (TA), and maturity index TSS/TA) remained unchanged between cold treated and control samples (Crifò et al. 2011).

In accessions of citron, limetta, sweet lime, lemon, and sweet orange, the acid-less phenotype is associated with large deletions or insertions of retro-transposons in the *Noemi* gene. In two accessions of limetta, a change in the core promoter region of *Noemi* is associated with reduced expression and increased pH of juice, indicating that *Noemi* is a major determinant of fruit acidity. *Noemi* in turn encodes a basic helix-loop-helix (*bHLH*) transcription factor and which controls flavonoid production as well as fruit acidity. A parallel research with 33 Citrus varieties completely unable to produce anthocyanins further reveals that acidless varieties containing functional alleles of *Ruby*, a key regulatory *MYB* gene, is essential for anthocyanin production. The two previous investigation show that both *bHLH* and *Ruby* are required for anthocyanins production. Mutation in any of the genes may affect pigmentation. *Noemi* encodes the *bHLH* protein that interacts with Ruby to control anthocyanin production in Citrus. *Noemi* and *Ruby* are both pleiotropic genes and play a regulatory role in anthocyanin production and fruit acidity (Butelli et al. 2019).

The biologists are searching for natural resistance in plants that could be a significant defense weapon for controlling biotic stresses in an environmental sustainable manner. A functional genomics approach employing cDNA microarrays with pathogen-infested flavedo tissue of Citrus fruits detected upregulation of few genes. The most highly induced genes were related to the phenylpropanoid pathway, engrossing 29 most up-regulated genes in the flavedo. Out of the 20 genes, 13 were involved in either phenylpropanoid metabolism or coumarin biosynthesis, including 7 different O-methyltransferases, isoflavone reductase, hydroxycinnamoyl transferase, 2 leucoanthocyanidin dioxygenases, and 2 SRG1 proteins. Genes related to methionine and ethylene biosynthetic processes, such as 1-aminocyclopropane-1-carboxylic acid oxidase (ACO) and proteins related to defense and response to stress were also induced in the flavedo. *EFE* and *CsACO* genes were over-expressed in tissues with direct affinity to ethylene production (Ballester et al. 2011).

The multiple examples presented in previous paragraphs unravels that transcriptome sequencing in plant is an efficient way of mining of functional genes, development of genomic markers, detection of discriminatory secondary metabolites, and analysis of related pathways (Guo et al. 2020). The advantage of RNA sequencing is that it provides distinctly different results owing to diverse developmental stages and tissues. Moreover, differential expression of genetic and epigenetic factors may be precisely estimated to provide information on diverse nutraceuticals production in Citrus and its valorization for industrial utilization (Table 3).

5.3 Nutra-metabolomics of Citrus

The quality of the Citrus fruit is directly related to the metabolic profile expressed by the plant. The primary and secondary metabolite deposition regulates the quality, taste, color, texture, flavor, appearance, and most importantly, the disease prevention property of the fruit. The color differences between Citrus varieties were associated with the carotenoid content. The Citrus color Index (CCI) value shows a significant positive correlation with the carotenoid content. Carotenoids are indispensable

Table 3 Important genes regulating nutraceutical production in Citrus

S. No	Gene ID	Gene	Nutraceutical	Source	Reference
1.	LOC102619013	*AN1* (*bHLH*), *AN2* (*MYB*), and *AN11* (WD-repeat)	Anthocyanin	Blood orange	Butelli et al. (2012)
2.		*CsMADS6*	Carotenoid	Sweet orange	Lu et al. (2018)
3.	LOC102630581	Limonoid UDP-glucosyltransferase	Limonoid	Sweet orange	Jia et al. (2019)
4.	LOC102608560	*F-box protein SKIP23-like*	Ascorbic acid	Sweet orange	Pillitteri et al. 2004
5.	LOC112098231	*Zeta-carotene desaturase*	Carotene development	Clementine	Terol et al. 2019
6.	LOC112100835	*Flavonoid 3′,5′-hydroxylase-like*	Flavonoid	Clementine	Itoh et al. 2016
7.	LOC102607309	*Chalcone synthase*	Flavonoid	Sweet orange	Wan et al. 2022
8.	Cstps1	*Sesquiterpene synthase*	Citral	Sweet orange	Sharon-Asa et al. 2003
9.	LOC112101189	*Coumarin 8-geranyltransferase 1b*	Coumarin	Clementine	Zhu et al. 2022
10.	LOC102625110	*Solanesyl diphosphate synthase 3*	Limonene	Clementine	Liu et al. 2020
11.	CtgOMT1	*O-methyltransferase (OMT)*	Polymethoxyflavones	Citrus grandis	Xian et al. 2022
12.	LOC102578043	*Citrus sucrose transporter 1*	Tangeritin	Sweet orange	Rooprai et al. 2021
13.	LOC18044720	*Squamosa promoter-binding-like protein 2*	Clementine	*Citrus clementina*	Zeng et al. 2019
14.	LOC102578043	*Citrus sucrose transporter 1*	Kutenone	Sweet orange	Hussain et al. 2020
15.	LOC102617171	*Fluoride export protein 2*	Camphor	Sweet orange	Li et al. 2021

molecules, providing protection against free radicals and oxidative stress (Zheng et al. 2019). Among organic acids, citric acid was the main organic acid found in Citrus flesh, followed by malic acid, acetic acid, and vitamin C. In a study of Korean Citrus varieties, significant variation was noticed among the local fruits (Hussain et al. 2017) with respect to acid content. Jeramon citrus showed the highest citric acid (482 mg/g DM) and malic acid content (60 mg/g DM), which were more or less 7–17 times and 2–9 times higher, respectively, than the other Korean varieties. Jeramon variety was developed from the nucellar embryo of a lemon plant, which excessively accumulates citric acid during fruit development. Kanpei (28 mg/g DM) and Natsumi (15 mg/g DM) had the lowest citric acid contents, while the lowest malic acid content was observed in Satsuma mandarin, Navel orange, Kanpei, and Setoka (approximately 6–8 mg/g DM). Setoka and Kanpei had the highest acetic acid content (about 6 mg/g DM), and the acetic acid content of the other varieties was approximately 2–4.8 mg/g DM. In particular, a high content of vitamin C, which has the strongest antioxidant activity in Citrus, was observed in Setoka (5.1 mg/g DM) and Jeramon (4.3 mg/g DM). These phenotypic and biochemical examinations provide a ground data for selection of germplasms and incorporation of the same for genomic and transcriptomic analysis, leading to the tracking of relevant biochemical pathways (Kim et al. 2021). The ratio of sugar content and organic acids is the main determinant of the maturity and core taste parameters in Citrus fruits. The sugar content and organic acid show a negative correlation, and the acid content declines during fruit development, resulting in a sweet taste in Citrus fruits (Smirnoff 2018).

A very interesting study elucidated the relationships between Citrus genotypes, diet, and health requirements of human subjects. Urinary metabolomic profile for volunteers identified proline betaine and flavanone glucuronides as known biomarkers, irrespective of Citrus genotypes. The study by the French research group revealed two strong discriminates identified as limonene 531 8,9-diol glucuronide and nootkatone 13,14-diol glucuronide belonging to terpene metabolite family in addition to standard markers (Pujos-Guillot et al. 2013). The unique study involving human subjects has thrown light on the recovery of nutritional biomarkers for accurate dietary assessment. Metabolome profiling could be a method to detect adulteration in fruit processing industry. The maintenance of the authenticity of food products is essential for export business. The compliance of end-products of our food industry with the international standard may lead to economic gain for the overall agro-business sector. Targeted and untargeted metabolomics could be applied for qualitative classification of authentic and adulterated samples.

According to the European Commission (2009), the addition of non-sweet orange (*Citrus sinensis*) to sweet orange juice is not allowed in the European Union countries. Codex Alimentarius guidelines state that up to 10% *orange* (*Citrus reticulata*) juice may be permitted in *Citrus sinensis* juice (Codex Alimentarius Commission 1992), while the Food and Drug Administration (FDA) permits the addition of 10% *Citrus sinensis* to pasteurized and canned orange juice, and up to 5% of *Citrus aurantium* to frozen concentrated orange juice. In proper operation of fruit industries, the assessment of adulteration is an important parameter. In this context, a scientific study led by the Indian food processing sector deserves mentioning.

Kinnow mandarin (*Citrus nobilis* × *Citrus deliciosa*), Jaffa, Mosambi orange (*Citrus sinensis*), and red blush grapefruit (*Citrus paradisi*) were collected from the Indian Agriculture Research Institute (IARI) for assessment of the authenticity of randomly collected samples with the authentic collection. Untargeted methods of ultra-performance liquid chromatography-quadrupole-time of flight mass spectrometry were applied to identify characteristic markers that could potentially be used to control Citrus fruit authenticity. The most influential markers identified were: didymin, rhoifolin, isorhoifolin, neohesperidin, hesperidin, naringin, narirutin, limonin glucoside, and vicenin-2. A targeted liquid chromatography-tandem mass spectrometry method was then optimized for the application of the identified markers. Diverse ratios among the identified phyto-chemicals act as potential biomarker for the *Citrus* juice industry. The tested biomarkers were proved authentic, with a proven record of reduction in adultery down to 2%. The study prescribed an untargeted qualitative approach with Principal Component Analysis (PCA) as a validation method for possible discrimination between authentic and adulterated samples (Jandrić et al. 2017). These integrative omics studies could accelerate the growth of nutraceutical industry and authentic biomarking of the Citrus sub-foods. The peel and seed waste of Citrus could lead to the development of a new industry with a simultaneous reduction in the global load of agro-waste.

6 Citrus Genome Database

The draft genome sequences of some *Citrus* species were available in the genomic databases of the Citrus Annotation Project (CAP) 3 and Phytozome. The bulk data of reference genomes from Citrus re-sequencing projects were available for application in population genetics, including genome-wide association studies (GWAS), evolutionary studies, and comparative genomics (Table 4). The genomic variation assists in the discovery of key quantitative trait loci (QTLs), molecular genetic markers, and genes relevant to important traits and contributes to the understanding of the origin and evolutionary relationships in Citrus. CitGVD (http://citgvd.cric.cn/home), a comprehensive database of Citrus genomic variations that provides a publicly available and free data service for scientific studies includes large sets of data on genomic variations (SNPs and INDELs) compiled from two released reference genomes for *Citrus clementina* and *Citrus grandis*, including 84 phenotypes, gene functional annotations, and informative literature. CitGVD also provides in-depth analysis, including CitTRAIT for phenotypic data statistics, CitGWAS for GWASs based on built-in data, CitEVOL for genetic evolution analysis, PCR primer design, and Gbrowse for variations and genes (Li et al. 2020).

Another Citrus Genome Database, known as CGD, is a USDA and NSF-funded resource that enables basic, translational, and applied research in Citrus (Dorrie and Sook 2018). It houses genomics, genetics, and breeding data for *Citrus* species. It is an open-source, generic database constructed on Tripal platform. The Citrus genome database contains 10 species: *C. clementina*, *C. ichangensis*, *C. sinensis*, *C. reticulata*, *C. medica*, *C. Maxima*, *C. limon*, *C. trifoliate*, *Atalantia buxifolia*,

Table 4 Brief genomic information of commercial Citrus species

Name	*C. sinensis*	*C. reticulata*	*C. limon*	*C. clementina*	*C. ichangensis*	*C. maxima*	*C. grandis*	*Poncirus Trifoliate*
Ploidy	Diploid	Diploid	Diploid	Diploid	Diploid	Diploid	Diploid	Diploid
Chromosome number	18	18	18	18	18	18	18	18
Genome size (Mb)	380	370	312	370	391	380	407	265
Available marker	2191	607	7	1968	0	8009	34	782
Available map	3	5	0	3	0	9	0	8
Available QTL	673	673	673	673	673	673	673	673
Available MTL	0	0	0	0	0	0	0	0
Available trait	229	229	229	229	229	229	229	229
Available genome	–	1	1	1	–	–	1	–

Fig. 3 Canopy structure of *Citrus reticulata* plant taken in the orchards of Darjeeling, West Bengal, India

Fortunella hindsii, and three species of *Ca. Liberibacter* (Fig. 3). It contains 366,169 genes and 649,803 mRNAs, 25 genome assemblies, 2255 germplasm, 85 maps, 60,407 markers, 16,971 phenotypic measurements, 75 trait descriptors, 6997 publications, and 673 QTLs for 153 agronomic traits. Seven tools are available such as BLAST, CitrusCyc, JBrowse, BIMS, Map Viewer, Synteny Viewer, and Expression Heatmap in the database for in-depth study of citrus genomes.

Another contemporary platform, CitSATdb (http://bioinfo.usu.edu/citSATdb/), mostly focuses on molecular markers mined from six *Citrus* species. The database is most useful for marker assisted selection and cisgenic improvement of Citrus. Recently, an updated genome information of *Citrus sinensis* (Wan et al. 2022) and 12 new sequenced genomes were integrated to provide an all-in-one database. The published Citrus genomes including Clementine, mandarin, pummelo, Mangshan wild mandarin, citron, Ichang papeda, kumquat, Trifoliate orange, and Chinese box orange were amalgamated to construct a more comprehensive database named the Citrus Pan-genome to Breeding Database (CPBD). CPBD presents large-scale datasets of Citrus transcriptomes, genome variations, and DNA methylomes as well as practical tools for Citrus breeding. In addition to omic data tool, some new datasets for CRISPR, KEGG/GO Enrichment, and GWAS are available in the CPBD platform (Liu et al. 2022b).

7 Nutraceutical Breeding for Designer Food Development in Citrus

In *Citrus*, several internal genetic complexities along with environment create hindrance in the improvement of different traits through breeding. The occurrence of heterozygosity, sexual reproduction, apomixis, polyploidy, juvenility, as well as recalcitrant seeds makes the breeding effort less productive. The majority of the Citrus plants are vegetatively propagated with preferential selection of the polyploid

types for larger fruit size with smooth rind. The traditional *Citrus* breeding mostly depended on selection of new cultivars from wild germplasm reserve and their domestication and cultivation. In *Citrus*, most of the agriculturally important traits are polygenic in nature, and conventional breeding has developed very few cultivars and rootstocks till date (Fig. 4). The scarcity of monogenic traits in *Citrus* is responsible for the lack of improvement of commercial traits. It has been noticed that a product of hybridization is only producing apomictic seeds, often diluting the effect of hybridization. Traditional breeding sometimes produces a hybrid with weak zygotic embryo as a result of inbreeding depression through mating of near-isogenic parents. The vast juvenile stage of the plants makes *Citrus* breeding a more time-intensive, resource draining, and land-occupying venture. In *Citrus*, conventional breeding were applied in dichotomous way for the improvement of both scions and rootstocks. In selected cases, the application of mutation breeding is noticed with gamma rays and other chemical mutagens.

In scion breeding, the new cultivars are developed through controlled crosses, and later the superior selections were grafted to compatible rootstocks. Natural hybridization, chance mutation, and Somaclonal variation have played major roles in cultivar development in *Citrus*. 'Satsuma' and 'Clementine' mandarins are outcomes of bud sport mutation. In grapefruit, a commercially acceptable cultivar, 'Star Ruby', was developed from 'Hudson' through irradiation. In Citrus, the long juvenile period was effectively utilized with a significant synergistic effect in the production of

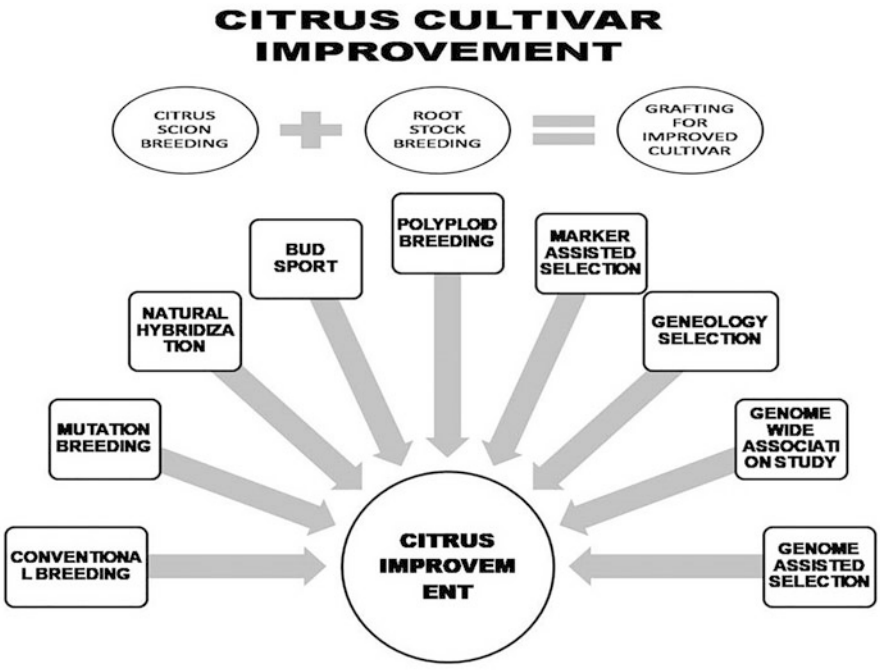


Fig. 4 Multiple breeding techniques employed for improvement of Citrus Industry

irradiated cultivars from growing scions. In sweet orange, lemon, and mandarin new varieties, 'Zhongyu No 7' (seedless), 'Kutdiken lemon' (resistant to *mal secco*), and 'Tango' (seedless) were developed respectively through the application of chemical mutagen in the juvenile stage of some scions. In scion breeding, there is a preference for monoembryonic species, inversely the rootstock propagation mostly depends on polyembryonic species, securing recovery of huge number of uniform plant-types. Root-stocks developed through inter-specific hybridization are found effective in disease resistance. Carrizo, Troyer Citranges, and Swingle Citrumelo were found to show resistance toward *Phytophthora* and nematode. Inter-specific hybrids between two species were found very effective for a third species. A hybrid between *Citrus reticulata* and *Poncirus trifoliata* was found exceptionally useful for *Citrus sinensis*. US-852, X639 are important representative of successful grafts involving three species. US early pride, UF Glow, Tango, Sugar belle, Roe tangerine, RBB7-34, Mandalate, Gold nugget, and Bingo, 950, 914 are commercially available varieties developed by the NVDMC, Citrus Research & Development Foundation (Gmitter et al. 2009). NRCC Mandarin Seedless-4, NRCC Acid Lime-7, NRCC Pummelo-5, NRCC Grapefruit-6, Cutter Valencia, Flame Grapefruit, US Pummelo-145, and Alemow were released by the Central Citrus Research Center, Nagpur, India, from the 552 indigenous and 62 core collections (Vi QRT Report-CCRI, 2018).

In agriculture, molecular marker-assisted breeding plays an effective role in improvement of crop quality and productivity. The mainstream agriculture has provided food security to the majority of global population. In the present era, a metabolomics-driven breeding could pave a way for the development of customized designer Citrus using system biology tools. The cross-talk among different biochemical pathways could produce varied nutrient elements. Application of several stress situations alters the production of nutrients in plants utilizing the same pathway. The previous discussion illuminated the differential role of multiple putative genes and transcription factors in the production of the consumer-friendly end products. Though nutraceutical breeding is an emerging technique, but few examples are available that exhibit equivalence to this futuristic breeding approach.

A study conducted on the influence of the rootstock in the production of bioactive compounds in the peels of a Mediterranean *Citrus* tree showed significant variation in the deposition of secondary compounds with respect to rootstock germplasm. Different rootstocks (Cleopatra mandarin and Troyer citrange) influenced the variable deposition of hesperidin and narirutin flavonoids. Flavanone glycosides, β-cryptoxanthin and violaxanthin, and limonene were the most abundant flavonoid, carotenoid, and limonoid identified in the peel essential oil of the grafted plants. The distinct difference in the content of bioactive compounds for the different groups of *Citrus* was in agreement with the taxonomic distinction of the rootstocks. The research highlighted that both mandarin and other hybrid orange varieties showed influence of the rootstocks in the deposition of a sizable amount of bioactive constituents, but the quantitative accumulation of individual nutrients were determined by the scion. This study concluded that different breeding programs with new rootstocks could yield conclusive outcomes for enhancement and production of desired nutraceutical according to the needs of the industry (Cano and Bermejo 2011).

In another study, examining the influence of *Citrus* rootstocks (Carrizo citrange, C-35 citrange, and F-5) in bioactive compound production on Clementine scion revealed a complex interaction between the rootstock in critical production of bioactive compounds. The species reflects a particular flavanone glycoside pattern. Six flavonoids are detected in clementines cv. 'Clemenrubí' and 'Orogrós'. The interaction between Orogrós cv and FA-5 rootstock presented the highest amounts of Api-6,8-di-C-glc, Nar-7-O-rut, and Hes-7-O-r flavonoids. Scion variety C-35 with citrange rootstock produced the heaviest and larger fruits, while the rootstock FA-5 produced the final reap with the major flavonoid content. This interesting study presents a special provision of integration of several desired agronomic and bio-chemical traits in a single plant by customized grafting techniques (Legua et al. 2017).

The genealogy research in Citrus has proven very successful in the identification of suitable breeding parents. In a study involving tropical Citrus plants, some germplasm were identified as a potential reservoir of bioactive compounds (Fig. 5). *Citrus maxima*, *Citrus grandis* L. (Pomelo), *Citrus paradisi* Macfad (Grapefruit), *Citrus sinensis* (Orange), *Citrus macroptera* (Wild Orange), *Citrus reticulata* (Mandarin), *Citrus limon* (Lemon), and *Citrus medica* L. (Citron) exhibit a varied morpho-metabolomic profile. Additionally, analysis of *Melicoccus bijugatus* Jacq. (Spanish Lime) fruit peels presented pharmacological potential and novel therapeutic proficiency. The research recommended the natural cultivation of the non-commercial Citrus for foodomics and sustainable agroforestry (Chel-Guerrero et al. 2022).

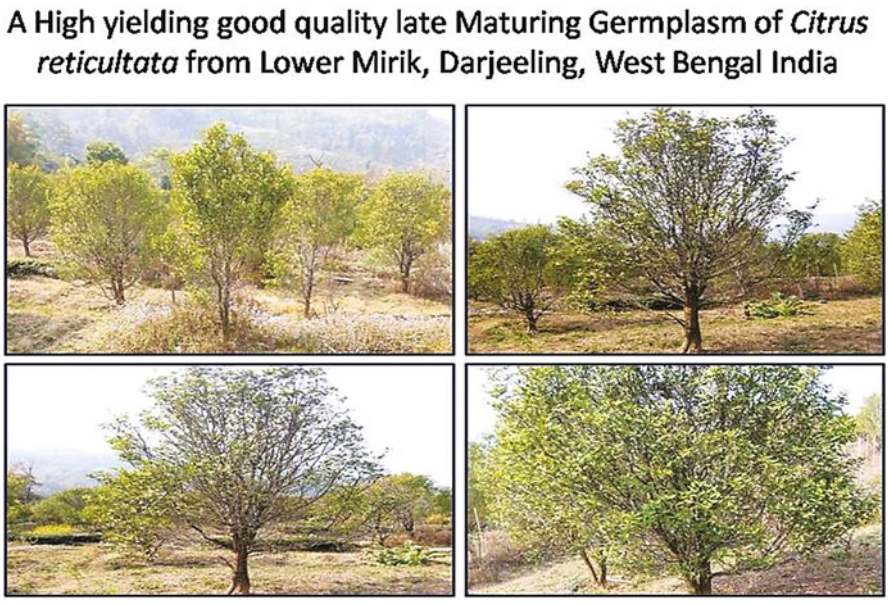


Fig. 5 Late maturing type Mandarin of Lower Mirik region of Darjeeling, West Bengal, India

The wild and endemic Citrus found in different parts of the globe may contribute toward the improvement of Citrus nutraceutical research. The gas chromatography–olfactometry (GC-O) analysis of volatiles of a wild mandarin, Mangshanyegan (*Citrus nobilis* Lauriro), and subsequent comparison with volatile profile of four *Citrus* species, Kaopan pummelo (*Citrus grandis*), Eureka lemon (*Citrus limon*), Huangyanbendizao tangerine (*Citrus reticulata*), and Seike navel orange (*Citrus sinensis*) exhibits significant distinctions. Monoterpene hydrocarbons d-limonene (85.75%) and β-myrcene (10.89%) pre-dominated the total volatile fraction. The flavor dilution factors (FD) detected eight oxygenated compounds, including (Z)- and (E)-linalool oxides specific to Mangshanyegan. The combined results of GC-O, quantitative analysis, odor activity values (OAVs), and omission tests confirms the balsamic and floral aroma of Mangshanyegan is controlled by β-myrcene and (Z)- and (E)-linalool oxides (Liu et al. 2012). These results are in consistence with the prospect of Citrus waste in the aroma and flavor industry.

A striking discovery of a new wild Citrus species native to the Ryukyu island increases the prospect of nutraceutical research in Citrus. This new species collection with eight wild Okinawan accessions forms a separate cluster, disowning their association with all previously sequenced species of Citrus. The accessions include 'tanibuta' type Citrus that is genetically distinct from tachibana and shiikuwasha described by Tanaka. Among their differences, *C. ryukyuensis* is a sexual species that produces monoembryonic seeds, while tachibana and shiikuwasha, both produces polyembryonic (apomictic) seeds. The identification of *C. ryukyuensis* as a pure species (a distinct sexually reproducing population without admixture) with low genome-wide heterozygosity (0.2–0.3%) with zygotic (sexual) reproduction forms a promising separate accession suitable as a breeding parent (Wu et al. 2021).

In a very recent study conducted in Manipur, India, confirms the origin of *Citrus indica* in the Indo-Burma border region. A wild orange morphologically resembling *C. indica* was characterized using morpho-taxonomic identifiers of Citrus. Additionally, plant barcoding using three chloroplast regions (trnL-F, psbK-I, matK-5′trnK spacer) and one nuclear (ITS) region established the identity of the Manipuri wild Citrus species as *Citrus indica*. The Dailong Village of Manipur, inhabited by the Rongmei tribes, were associated with the natural conservation and maintenance of wild Citrus known as 'Garuan-thai'. *C. indica* is known for its poor regeneration and less adaptability to new habitats. A distinct hilly microclimate is essential for its survival and existence (Fig. 6). The identification of 'Garuan-thai' in North Eastern India unveils the prospect of discovering potent genotypes for nutraceutical exploitation (Devi et al. 2022).

In recent years, several attractive breeding strategies were applied for improvement of Citrus orchard and fruit production strategies. Normally, Citrus seedlings go through a very long juvenile phase (about 3–20 years), which hinders the breeding and improvement of Citrus. Research on minimizing the juvenile phase and promotion of early flowering in Citrus was done by adoption of a transgenic breeding method. APETALA1 (*AP1*) and LEAFY (*LFY*) genes were introduced into citrange (*P. trifoliata* L. Raf. × hybrid of *C. sinensis* L. Osbeck) from Arabidopsis. Transgenic citrange plants over-expressing the *AP1* and *LFY* genes flowered early and

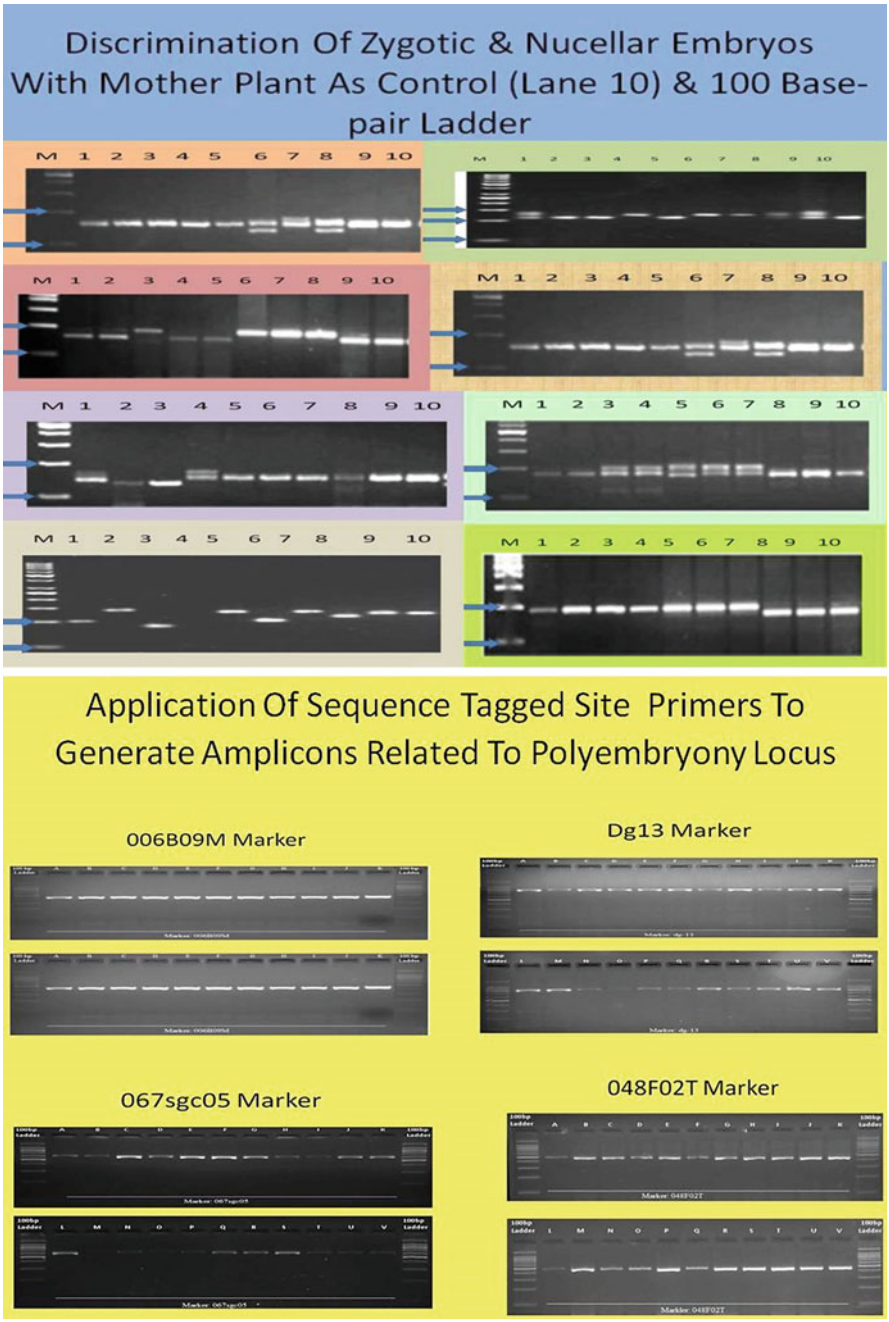


Fig. 6 Molecular Profile of some elite clones of mandarin (*Citrus reticulata*) orange collected from North Eastern Himalayan Region of Indian Subcontinent

produced fertile and normal flowers (Pillitteri et al. 2004). Fruits were obtained from first-year transgenic plants that were only 2–20 months old and bore their first flowers, significantly reducing the juvenile phase to less than 5 years in transgenic citrange compared with control plants (Peña et al. 2001). Zygotic seedlings obtained by crossing *AP1* and *LFY*-transgenic citrange also showed early flowering with normal fruit setting in the first spring, validating provisions for novel research (Rao et al. 2021). The *Citrus* genetic resource in conjecture with model and non-model plants could produce unique nutraceutical cultivar in *Citrus*. The gene co-suppression, feedback inhibition regulates metabolic inheritance in *Citrus* and transcriptomic tool could encourage anabolic biosynthesis of desired components.

In new generation *Citrus* breeding, a novel approach of genomic-assisted breeding (GAB) plays an important role. This GAB method could solve the main three constraints of Citrus improvement, meeting the commercial demand, disease resistance, and fast breeding. The GAB technique is an integration of genome wide association studies (GWAS), and genome selection (GS) is expected to increase the prediction accuracy of conventional MAS. In this method, the parents of multiple Citrus cultivars mainly natural hybrids will be confirmed. The genomic pedigree will be beneficial for selection of some unfamiliar cultivar for hybridization to assimilate novel traits in hybrid by reduction in unwanted traits. A research conducted by NARO released the current cultivars 'Aurastar', Nou No. 7, 'Nou No. 8', second-generation Trifoliate orange from a cross of 'TF Flying Dragon' and a local cultivar 'Hassaku', followed by subsequent crossing with 'Kiyomi', for transfer of Citrus tristeza resistance trait from the Banpeiyu pummelo genotype (Goto et al. 2018). The genealogy study of the cultivars is a productive approach that could accelerate the breeding pace. Several Japanese cultivars Satsuma, Hassaku, Sudachi, and Kabosu having unique traits were selected within a single cross. Elite hybrids were selected from the crossing of 'Kaikoukan' and 'tachibana' within a single generation. Screening of thousands and tens of thousands of offspring combining GAB techniques was assumed to be the new savior of dwindling orchards and the Citrus industry (Shimizu 2022).

8 Conclusion

In current times, the global nutrition sector is facing a dichotomous constraint: one is the nutritional deficit and malnutrition, and the other is anomalous nutrition associated with luxuriance in food habits. The malnutrition of children due to lack of healthy diet in developing and some developed nations is recognized as a global problem, the other condition of anomalous nutrition promotes obesity, diabetes, and coronary heart diseases in developed countries with abundance availability of high-calorie food. The imbalance in nutritional requirement could be compensated by the intake of functional foods. In tropical and subtropical countries, the fruits and vegetables are loaded with diverse nutraceuticals. A vast majority of economically

backward countries are blessed with rich biodiversity reserves. This rich plant gene-pool is a robust weapon for the countries to alleviate poverty by creation of domestic and global agri-business. In real life, these germplasm biomass produces a sizable agro-waste that becomes a concern for the under-privileged nations. Through an intelligent approach, these agro-wastes could be utilized in alleviation of nutritional inconsistencies. Citrus is one of the prominent fruit crop with immense potential to be included as raw material in the emerging nutraceutical sector. This fruit crop has a long history of cultivation and has conserved a vast wild gene reserve in tropical and sub-tropical regions of the world. The appropriate utilization of the omic tool could lead to metabolic flux of the nutritive and pharmaceutically valuable plant bio-actives. Genotype-driven measurable phenotypic changes with consequent understanding of the biochemical mechanism and pathway dynamics may assist in food-omics research. The cutting-edge technologies of new omic era could increase the metabolite heritability, leading to development of designer Citrus cultivar. Omic-guided novel designer citrus may accelerate the pace of emerging nutraceutical industry and ensure intelligent agro-waste management.

References

Abeynayake SW, Panter S, Chapman R, Webster T, Rochfort S, Mouradov A, Spangenberg G (2012) Biosynthesis of proanthocyanidins in white clover flowers: cross talk within the flavonoid pathway. Plant Physiol 158(2):666–678

Babarinde SA, Kemabonta A, Olatunde OZ, Ojutiku EO, Adeniyi A (2021) Composition and toxicity of rough lemon (*Citrus jambhiri* Lush.) rind essential oil against red flour beetle. Acta Ecol Sin 41(4):325–331. https://doi.org/10.1016/j.chnaes.2020.07.001

Ballester AR, Lafuente MT, Forment J, Gadea J, De Vos RC, Bovy AG, Gonzalez-Candelas LU (2011) Transcriptomic profiling of Citrus fruit peel tissues reveals fundamental effects of phenylpropanoids and ethylene on induced resistance. Mol Plant Pathol 12(9):879–897

Brito A, Ramirez JE, Areche C, Sepúlveda B, Simirgiotis MJ (2014) HPLC-UV-MS profiles of phenolic compounds and antioxidant activity of fruits from three Citrus species consumed in Northern Chile. Molecules 19(11):17400–17421

Brophy JJ, Goldsack RJ, Paul FI (2001) The leaf oils of the Australian species of Citrus (Rutaceae). J Essent Oil Res 13(4):264–268. https://doi.org/10.1080/10412905.2001.9699690

Butelli E, Licciardello C, Zhang Y, Liu J, Mackay S, Bailey P, Reforgiato-Recupero G, Martin C (2012) Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. Plant Cell 24:1242–1255

Butelli E, Licciardello C, Ramadugu C, Durand-Hulak M, Celant A, Recupero GR, Froelicher Y, Martin C (2019) Noemi controls production of flavonoid pigments and fruit acidity and illustrates the domestication routes of modern Citrus varieties. Curr Biol 29(1):158–164

Cano A, Bermejo A (2011) Influence of rootstock and cultivar on bioactive compounds in Citrus peels. J Sci Food Agric 91:1702–1711. https://doi.org/10.1002/jsfa.4375

Celano R, Campone L, Pagano I, Carabetta S, Di Sanzo R, Rastrelli L, Piccinelli AL, Russo M (2019) Characterisation of nutraceutical compounds from different parts of particular species of *Citrus sinensis* 'Ovale Calabrese' by UHPLC-UV-ESI-HRMS. Nat Prod Res 33(2):244–251

Chel-Guerrero LD, Cuevas-Glory LF, Sauri-Duch E, Sierra-Palacios E, De Leon-Sanchez FD, Mendoza-Espinoza JA (2022) Tropical fruit peels as sources of bioactive compounds: a review. Pak J Bot 54(3):1169–1179

Chhikara N, Kour R, Jaglan S, Gupta P, Gata Y, Panghal A (2018) *Citrus medica*: nutritional, phytochemical composition and health benefits – a review. Food Funct 9:1978–1992. https://doi.org/10.1039/C7FO02035J

Codex Alimentarius Commission (1992) Fruit juices and related products, Rome. Commission Directive 2009/106/EC amending Commission Directive 2001/112/EC relating to fruit juices and certain similar products intended to human consumption

Crifò T, Puglisi I, Petrone G, Recupero GR, Piero AR (2011) Expression analysis in response to low temperature stress in blood oranges: implication of the flavonoid biosynthetic pathway. Gene 1: 476(1–2):1–9

Das L, Bhaumik E, Raychaudhuri U, Chakraborty R (2012) Role of nutraceuticals in human health. J Food Sci Technol 49(2):173–183. https://doi.org/10.1007/s13197-011-0269-4

Devi EJ, Labala RK, Sanabam R, Singh NS, Modak R, Devi HS (2022) New report of *Citrus indica* Yu. Tanaka, a wild progenitor species of Citrus from Dailong Forest, Manipur, and recommendation for its conservation. Genet Resour Crop Evol 69(2):545–558

Dorrie S, Sook J (2018) Citrus Genome Database. MainLab Bioinformatics, Washington State University. https://data.nal.usda.gov/dataset/Citrus-genome-database. Accessed 28 Nov 2022. (dataset) Main

Dutta F, Marak MC, Meena AK (2017) Nutraceuticals from major fruit crops. Int J Agric Innov Res 5(6):2319–1473

Eneke IC, Essien EB, Wegu MO (2021) Gas chromatography-mass spectrometry (GC-MS) analysis of bioactive components present in grape Citrus peel in Nigeria. GSC Adv Res Rev 8(01):166–174. https://doi.org/10.30574/gscarr.2021.8.1.0152

European Union Law, *Regulation (EC) No 810/2009 of the European Parliament and of the Council of 13 July 2009 establishing a Community Code on Visas (Visa Code)*, 2009, Document 32009R0810 (2009).

Gmitter FG, Soneji JR, Rao MN (2009) Citrus breeding. In: Breeding plantation tree crops: temperate species. Springer, New York, pp 105–134

Goto S, Yoshioka T, Ohta S, Kita M, Hamada H et al (2018) QTL mapping of male sterility and transmission pattern in progeny of Satsuma mandarin. PLoS One 13(7):e0200844. https://doi.org/10.1371/journal.pone.0200844

Guo L, Liu Y, Luo L, Hussain SB, Bai Y, Alam SM (2020) Comparative metabolites and citrate-degrading enzymes activities in Citrus fruits reveal the role of balance between ACL and Cyt-ACO in metabolite conversions. Plan Theory 9(3):350. https://doi.org/10.3390/plants9030350

Hamdan DI, Abdulla RH, Mohamed ME, El-Shazly AM (2013) Chemical composition and biological activity of essential oils of Cleopatra mandarin (*Citrus reshni*) cultivated in Egypt. J Pharmacognos Phytotherap 5(5):83–90. https://doi.org/10.5897/JPP2013.0277

Harsha H, Aarti S (2015) Quality evaluation of herbal juice developed from traditional Indian medicinal plants using *Citrus limetta* as base. J Nutr Food Sci 5:5. https://doi.org/10.4172/2155-9600.1000396

Herman Z, Hasegawa S, Fong CH, Ou P (1989) Limonoids in *Citrus ichangensis*. J Agric Food Chem 37(4):850–851

Hojjati M, Barzegar H (2017) Chemical composition and biological activities of lemon (*Citrus limon*) leaf essential oil. Nutr Food Sci Res 4:15–24. https://doi.org/10.29252/nfsr.4.4.3

Hussain S, Shi CY, Guo LX, Kamran HM, Sadka A, Liu YZ (2017) Recent advances in the regulation of citric acid metabolism in Citrus fruit. Crit Rev Plant Sci 36:241–256

Hussain SB, Guo LX, Shi CY, Khan MA, Bai YX, Du W, Liu YZ (2020) Assessment of sugar and sugar accumulation-related gene expression profiles reveal new insight into the formation of low sugar accumulation trait in a sweet orange (*Citrus sinensis*) bud mutant. Mol Biol Rep 47(4): 2781–2791. https://doi.org/10.1007/s11033-020-05387-6

Ito C, Fujiwara K, Kajita M, Ju-Ichi M, Takemura Y, Suzuki Y, Tanaka K, Omura M, Furukawa H (1991) New Coumarins from Citrus plants. Chem Pharmaceut Bull 39(10):2509–2513. https://doi.org/10.1248/cpb.39.2509

Itoh N, Iwata C, Toda H (2016) Molecular cloning and characterization of a flavonoid-*O*-methyltransferase with broad substrate specificity and regioselectivity from *Citrus depressa*. BMC Plant Biol 16:180. https://doi.org/10.1186/s12870-016-0870-9

Jabri KI, Marzouk B (2013) Characterization of bioactive compounds in Tunisian bitter orange (*Citrus aurantium* L.) peel and juice and determination of their antioxidant activities. Biomed Res Int. https://doi.org/10.1155/2013/345415

Jandrić Z, Islam M, Singh DK, Cannavan A (2017) Authentication of Indian Citrus fruit/fruit juices by untargeted and targeted metabolomics. Food Control 72:181–188

Jia N, Liu J, Tan P, Sun Y, Lv Y, Liu J, Sun J, Huang Y, Lu J, Jin N, Li M, Md Sharif Uddin Imam K, Xin F, Fan B (2019) *Citrus sinensis* MYB transcription factor CsMYB85 induce fruit juice sac lignification through interaction with other CsMYB transcription factors. Front Plant Sci 10: 213. https://doi.org/10.3389/fpls.2019.00213

Khan UM, Sameen A, Aadil RM, Shahid M, Sezen S, Zarrabi A, Ozdemir B, Sevindik M, Kaplan DN, Selamoglu Z, Ydyrys A, Anitha T, Kumar M, Sharifi-Rad J, Butnariu M (2021) *Citrus* genus and its waste utilization: a review on health-promoting activities and industrial application. J Evid Based Integr Med 2021:2488804. https://doi.org/10.1155/2021/2488804

Kim D-S, Lee S, Park SM, Yun SH, Gab H-S, Kim SS, Kim H-J (2021) Comparative metabolomics analysis of Citrus varieties. Foods 10(11):2826. https://doi.org/10.3390/foods10112826

Legua P, Hernández F, Ángeles M (2017) Influence of Citrus rootstocks in bioactive compounds of clementines. J Food Nutr Res 5(8):545–552

Li Q, Qi J, Qin X et al (2020) CitGVD: a comprehensive database of Citrus genomic variations. Hortic Res 7:12. https://doi.org/10.1038/s41438-019-0234-3

Li B, Zhang Y, Qiu D, Francis F, Wang S (2021) Comparative proteomic analysis of sweet orange petiole provides insights into the development of Huanglongbing symptoms. Front Plant Sci 12: 656997

Liu C, Cheng Y, Zhang H, Deng X, Chen F, Xu J (2012) Volatile constituents of wild Citrus Mangshanyegan (*Citrus nobilis* Lauriro) peel oil. J Agric Food Chem 60(10):2617–2628

Liu M, Liu X, Hu J, Xue Y, Zhao X (2020) Genetic diversity of limonene synthase genes in Rongan kumquat (*Fortunella crassifolia*). Funct Plant Biol 47(5):425–439. https://doi.org/10.1071/FP19051

Liu S, Li Y, Fang H, Huang B, Zhao C, Sun C, Li S, Chen K (2022a) Genome-wide identification and expression analysis of MATE gene family in Citrus fruit (*Citrus clementina*). Genomics 114 (5):110446

Liu H, Wang X, Liu S, Qamar MTU, Xu Q, Chen L-L (2022b) Citrus Pan-Genome to Breeding Database (CPBD): a comprehensive genome database for Citrus breeding. https://doi.org/10.1016/j.molp.2022.08.006

Loizzo MR, Leporini M, Sicari V et al (2018) Investigating the in vitro hypoglycaemic and antioxidant properties of *Citrus* × *clementina* Hort. juice. Eur Food Res Technol 244:523–534. https://doi.org/10.1007/s00217-017-2978-z

Lota ML, de Rocca Serra D, Tomi F, Bessiere JM, Casanova J (1999) Chemical composition of peel and leaf essential oils of *Citrus medica* L. and *C. limonimedica* Lush. Flav Frag J 14:161–166. https://doi.org/10.1002/(SICI)1099-1026(199905/06)14:3<161::AID-FFJ801>3.0.CO;2-8

Lu S, Zhang Y, Zhu K, Yang W, Ye J, Chai L, Xu Q, Deng X (2018) The Citrus transcription factor CsMADS6 modulates carotenoid metabolism by directly regulating carotenogenic genes. Plant Physiol 176:2657–2676

Luan YT, Wan SS, Wang RQ, Xu CJ (2020) Accumulation of red apocarotenoid beta-citraurin in peel of a spontaneous mutant of huyou (*Citrus changshanensis*) and the effects of storage temperature and ethylene application. Food Chem 309:125705

Mahato N, Sharma K, Sinha M, Choa MH (2018) Citrus waste derived nutra-pharmaceuticals for health benefits: current trends and future perspectives. J Funct Foods 40:307–316. https://doi.org/10.1016/j.jff.2017.11.015

Medina-Torres N, Espinosa-Andrews H, Trombotto S, Ayora-Talavera T, Patrón-Vázquez J, González-Flores T, Sánchez-Contreras Á, Cuevas-Bernardino JC, Pacheco N (2019)

Ultrasound-assisted extraction optimization of phenolic compounds from *Citrus latifolia* waste for chitosan bioactive nanoparticles development. Molecules 24(19):3541. https://doi.org/10.3390/molecules24193541

Mondal B (2021) Application of molecular marker in effective management of mandarin orange orchards of West Bengal. ResJ Agric Sci 12(1):126–130

National Center for Biotechnology Information (NCBI), *Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988]*, 2023, KA-03391 (2023).

National Center for Biotechnology Information (NCBI) [Internet]. National Library of Medicine (US), National Center for Biotechnology Information, Bethesda. 1988 [cited 2022 Nov 6]. Available from: https://www.ncbi.nlm.nih.gov/

Navarra M, Mannucci C, Delbò M, Calapai G (2015) *Citrus bergamia* essential oil: from basic research to clinical application. Front Pharmacol Ethnopharmacol. https://doi.org/10.3389/fphar.2015.00036

Nishad J, Singh SP, Singh S, Saha S, Dubey AK, Varghese E, Kaur C (2018) Bioactive compounds and antioxidant activity of selected Indian pummelo (*Citrus grandis* L. Osbeck) germplasm. Sci Hortic 233:446–454

Oboh G, Ademosun AO (2011) Shaddock peels (*Citrus maxima*) phenolic extracts inhibit α-amylase, α-glucosidase and angiotensin I-converting enzyme activities: a nutraceutical approach to diabetes management. Diabetes Metab Syndr Clin Res Rev 5(3):148–152. https://doi.org/10.1016/j.dsx.2012.02.008

Pal RS, Pal Y, Singh V (2016) Isolation and chemical characterization of volatile constituents of leaves of *Citrus limon* (Linn.). Int J Pharmaceut Technol 8:25126–25132

Peña L, Martín-Trillo M, Juárez J, Pina JA, Navarro L, Martínez-Zapater JM (2001) Constitutive expression of Arabidopsis LEAFY or APETALA1 genes in Citrus reduces their generation time. Nat Biotechnol 19(3):263–267

Pikaard CS, Scheid OM (2014) Epigenetic regulation in plants. Cold Spring Harb Perspect Biol 6(12):a019315. https://doi.org/10.1101/cshperspect.a019315

Pillitteri LJ, Lovatt CJ, Walling LL (2004) Isolation and characterization of LEAFY and APETALA1 homologues from *Citrus sinensis* L. Osbeck 'Washington'. J Amer Soc Hortic Sci 129(6):846–856

Pujos-Guillot E, Hubert J, Martin JF, Lyan B, Quintana M, Claude S, Chabanas B, Rothwell JA, Bennetau-Pelissero C, Scalbert A, Comte B (2013) Mass spectrometry-based metabolomics for the discovery of biomarkers of fruit and vegetable intake: Citrus fruit as a case study. J Proteome Res 12(4):1645–1659

Rao MJ, Zuo H, Xu Q (2021) Genomic insights into Citrus domestication and its important agronomic traits. Plant Commun (1):100138. https://doi.org/10.1016/j.xplc.2020.100138

Raspo MA, Vignola MB, Andreatta AE, Juliani HR (2020) Antioxidant and antimicrobial activities of Citrus essential oils from Argentina and the United States. Food Biosci 36:100651

Rodríguez-Casado A (2014) The health potential of fruits and vegetables phytochemicals: notable examples. Crit Rev Food Sci Nutr:56. https://doi.org/10.1080/10408398.2012.755149

Rooprai HK, Christidou M, Murray SA, Davies D, Selway R, Gullan RW, Pilkington GJ (2021) Inhibition of invasion by polyphenols from Citrus fruit and berries in human malignant glioma cells in vitro. Anticancer Res 41(2):619–633. https://doi.org/10.21873/anticanres.14813

Rossi RC, da Rosa SR, Weimer P, Moura JGL, de Oliveira VR, de Castilhos J (2020) Assessment of compounds and cytotoxicity of *Citrus deliciosa* Tenore essential oils: from an underexploited by-product to a rich source of high-value bioactive compounds. Food Biosci 38:100779

Russo M, Arigò A, Calabrò ML, Farnetti S, Mondello L, Dugo P (2016) Bergamot (*Citrus bergamia* Risso) as a source of nutraceuticals: limonoids and flavonoids. J Funct Foods 20:10–19. https://doi.org/10.1016/j.jff.2015.10.005

Russo M, Bonaccorsi IL, Arigò A, Cacciola F, Gara LD, Dugo P, Mondello L (2021) Blood orange (*Citrus sinensis*) as a rich source of nutraceuticals: investigation of bioactive compounds in different parts of the fruit by HPLC-PDA/MS. Nat Prod Res 35(22):4606–4610. https://doi.org/10.1080/14786419.2019.1696329

Safdar MN, Kausar T, Jabbar S, Mumtaz A, Ahad K, Saddozai AA (2017) Extraction and quantification of polyphenols from kinnow (*Citrus reticulate* L.) peel using ultrasound and maceration techniques. J Food Drug Analys 25(3):488–500. https://doi.org/10.1016/j.jfda.2016.07.010

Saini RK, Keum YS (2018) Carotenoid extraction methods: a review of recent developments. Food Chem 240:90–103

Saini RK, Ranjit A, Sharma K, Prasad P, Shang X, Gowda KGM, Keum Y-S (2022) Bioactive compounds of Citrus fruits: a review of composition and health benefits of carotenoids, flavonoids, limonoids, and terpenes. Antioxidants 11(2):239. https://doi.org/10.3390/antiox11020239

Sanmartin C, Taglieri I, Macaluso M, Sgherri C, Ascrizzi R, Flamini G, Venturi F, Quartacci MF, Luro F, Curk F, Pistelli L (2019) Cold-pressing olive oil in the presence of cryomacerated leaves of olea or Citrus: nutraceutical and sensorial features. Molecules 24(14):2625

Scora RW, Ahmed M (1995) Essential leaf oil composition of *Eremocitrus glauca* (Lindl.) Swing., an Aurantioid xerohyts. J Essential Oil Res 7:579–584

Shaikh IA, Muddapur UM, Bagewadi ZK, Chiniwal S, Ghoneim MM, Mahnashi MH, Alsaikhan F, Yaraguppi D, Niyonzima FN, More SS, Mannasaheb BA, Al Ali A, Asiri A, Khan AA, Iqubal SMS (2022) Characterization of bioactive compounds from *Acacia concinna* and *Citrus limon*, silver nanoparticles' production by *A. concinna* extract, and their biological properties. Molecules 27(9):2715. https://doi.org/10.3390/molecules27092715

Sharon-Asa L, Shalit M, Frydman A, Bar E, Holland D, Or E, Lavi U, Lewinsohn E, Eyal Y (2003) Citrus fruit flavor and aroma biosynthesis: isolation, functional characterization, and developmental regulation of Cstps1, a key gene in the production of the sesquiterpene aroma compound valencene. Plant J 36:664–674. https://doi.org/10.1046/j.1365-313X.2003.01910.x

Shimizu T (2022) Breeding new premium quality cultivars by Citrus breeding 2.0 in Japan: an integrative approach suggested by genealogy. Horticulturae 8(6):559

Shon MY, Park SK (2006) Synergistic effect of Yuza (*Citrus junos*) extracts and ascorbic acid on antiproliferation of human cancer cells and antioxidant activity. Kor J Food Preserv:13

Silalahi J (2002) Anticancer and health protective properties of Citrus fruit components. Asia Pacif J Clin Nutr 11:79–84. https://doi.org/10.1046/j.1440-6047.2002.00271.x

Siti HN, Mohamed S, Kamisah Y (2022) Potential therapeutic effects of *Citrus hystrix* DC and its bioactive compounds on metabolic disorders. Pharmaceuticals 15(2):167. https://doi.org/10.3390/ph15020167

Smirnoff N (2018) Ascorbic acid metabolism and functions: a comparison of plants and mammals. Free Radic Biol Med 122:116–129

Strazzer P, Spelt CE, Li S, Bliek M, Federici CT, Roose ML, Koes R, Quattrocchio FM (2019) Hyperacidification of Citrus fruits by a vacuolar proton-pumping P-ATPase complex. Nat Commun 10:1–11

Sun Q, Zheng L, Wan-Peng X, Shao-Lan H, Shi-Lai Y, Qiang L, Yong-Qiang Z, Deng L, Rang-Jin X (2014) Transcriptome analysis of blood orange (*Citrus sinensis*) following fruit bagging treatment by digital gene expression profiling. J Hortic Sci Biotechnol 89(4):397–407

Terol J, Nueda M, Ventimilla D et al (2019) Transcriptomic analysis of *Citrus clementina* mandarin fruits maturation reveals a MADS-box transcription factor that might be involved in the regulation of earliness. BMC Plant Biol 19:47. https://doi.org/10.1186/s12870-019-1651-z

Testai L, Calderone V (2017) Nutraceutical value of Citrus flavanones and their implications in cardiovascular disease. Nutrients 9(5):502. https://doi.org/10.3390/nu9050502

Tocmo R, Pena-Fronteras J, Calumba KF, Mendoza M, Johnson JJ (2020) Valorization of pomelo (*Citrus grandis* Osbeck) peel: a review of current utilization, phytochemistry, bioactivities, and mechanisms of action. Comprehens Rev Food Sci Food Saf 19:1969–2012. https://doi.org/10.1111/1541-4337.12561

Tsitsagi M, Ebralidze K, Chkhaidze M, Rubashvili I, Tsitsishvili V (2018) Sequential extraction of bioactive compounds from tangerine (*Citrus Unshiu*) peel. Ann Agrar Sci 16(2):236–241. https://doi.org/10.1016/j.aasci.2018.02.007

Tundis R, Loizzo MR, Bonesi M, Menichini F, Mastellone V, Colica C, Menichini F (2012) Comparative study on the antioxidant capacity and cholinesterase inhibitory activity of *Citrus aurantifolia* Swingle, *C. aurantium* L., and *C. bergamia* Risso and Poit. peel essential oils. J Food Sci 77(1):H40–H46

Vekiari SA, Protopapadakis EE, Papadopoulou P, Papanicolaou D, Panou C, Vamvakias M (2002) Composition and seasonal variation of the essential oil from leaves and peel of a Cretan lemon variety. J Agric Food Chem 50:147–153. https://doi.org/10.1021/jf001369a

VI QRT Report (2013-2018), *ICAR- Central Citrus Research Institute*, 2018, QRT-CCRI, Nagpur 1 (2023).

Wan Q, Bai T, Liu M, Liu Y, Xie Y, Zhang T, Huang M, Zhang J (2022) Comparative analysis of the chalcone-flavanone isomerase genes in six Citrus species and their expression analysis in sweet orange (*Citrus sinensis*). Front Genet 13:848141. https://doi.org/10.3389/fgene

Wang F, Wang M, Liu X, Xu Y, Zhu S, Shen W, Zhao X (2017) Identification of putative genes involved in limonoids biosynthesis in Citrus by comparative transcriptomic analysis. Front Plant Sci 8:782

Wang Y, Ji S, Zang W, Wang N, Cao J, Li X, Sun C (2019) Identification of phenolic compounds from a unique citrus species, finger lime (*Citrus australasica*) and their inhibition of LPS-induced NO-releasing in BV-2 cell line. Food Chem Toxicol:129. https://doi.org/10.1016/j.fct.2019.04.006

Wu GA, Sugimoto C, Kinjo H et al (2021) Diversification of mandarin Citrus by hybrid speciation and apomixis. Nat Commun 12:4377. https://doi.org/10.1038/s41467-021-24653-0

Xian L, Sahu SK, Huang L, Fan Y, Lin J, Su J, Bai M, Chen Y, Wang S, Ye P, Wang F, Luo Q, Bai H, Lin X, Yuan C, Geng X, Liu H, Wu H (2022) The draft genome and multi-omics analyses reveal new insights into geo-herbalism properties of *Citrus grandis* 'Tomentosa'. Plant Sci 325: 111489. https://doi.org/10.1016/j.plantsci.2022.111489

Xiao H, Yang CS, Shiming L, Huanyu J, Ho C-T, Patel T (2009) Monodemethylated polymethoxyflavones from sweet orange (*Citrus sinensis*) peel inhibit growth of human lung cancer cells by apoptosis. Mol Nutr Food Res 53(3):398–406

Yu F, Xu X, Lin S, Peng T, Zeng S (2022) Integrated metabolomics and transcriptomics reveal flavonoids glycosylation difference in two Citrus peels. Sci Hortic 292:110623

Zeng RF, Zhou JJ, Liu SR, Gan ZM, Zhang JZ, Hu CG (2019) Genome-wide identification and characterization of SQUAMOSA-promoter-binding protein (SBP) genes involved in the flowering development of *Citrus clementina*. Biomol Ther 9(2):66. https://doi.org/10.3390/biom9020066

Zhao XJ, Xing TT, Li YF, Jiao BN, Jiang D (2018) Efficient analysis of phytochemical constituents in the peel of Chinese wild citrus Mangshanju (*Citrus reticulata* Blanco) by ultra high performance liquid chromatography–quadrupole time-of-flight-mass spectrometry. J Separ Sci 41(9): 1947–1959

Zheng X, Tang Y, Ye J, Pan Z, Tan M, Xie Z, Chai L, Xu Q, Fraser PD, Deng X (2019) SLAF-based construction of a high-density genetic map and its application in QTL mapping of carotenoids content in Citrus fruit. J Agric Food Chem 67:994–1002

Zhu C, Peng C, Qiu D, Zeng J (2022) Metabolic profiling and transcriptional analysis of carotenoid accumulation in a red-fleshed mutant of pummelo (*Citrus grandis*). Molecules 27(14):4595. https://doi.org/10.3390/molecules27144595

Watermelon: Advances in Genetics of Fruit Qualitative Traits

Sudip Kumar Dutta, Padma Nimmakayala, and Umesh K. Reddy

Contents

Abstract

Watermelon (*Citrullus lanatus*) belongs to the family Cucurbitaceae. The crop is grown commercially in regions with extended warm, frost-free months. Watermelon is cultivated for its colorful, tender, juicy, and sweet fruit. They are generally consumed fresh and make an excellent and delicious dessert, particularly during the summer months. Because of its smaller genome and large number of gene mutations, watermelon is a suitable crop species for genetic research. Watermelon's genome has 424 million base pairs. DNA sequencing found significant conservation, which is relevant for comparative genomics within Cucurbitaceae and other species.

S. K. Dutta (✉)
Gus R. Douglass Institute, Department of Biology, West Virginia State University, Institute, WV, USA

ICAR RC NEH Region, Sikkim centre, Gangtok, Sikkim, India

P. Nimmakayala · U. K. Reddy
Gus R. Douglass Institute, Department of Biology, West Virginia State University, Institute, WV, USA
e-mail: padma@wvstateu.edu; ureddy@wvstateu.edu

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_36

There exists a huge genetic variability in the fruit quality characteristics of watermelon with respect to seed traits, fruit shape, fruit size, skin color, skin pattern, flesh color and sugar/acid composition, fruit bitterness, and many more. This chapter serves as a guide to show the prospects and advances made in the genetics of fruit qualitative traits in watermelon breeding programs depending on profitability and consumer preferences.

Keywords

Watermelon · Fruit · Genomics · Quality · Flesh color · Rind pattern

1 Introduction

Watermelon (*Citrullus lanatus* [Thunb.] Matsum. and Nakai var. *lanatus*; 2n = 2x = 22) belonging to the family Cucurbitaceae is primarily cultivated for its fresh and nutritious fruit. *Citrullus lanatus* var. *citroide*s, *C. naudinianus, C. mucosospermus*, *C. rehmii*, *C. ecirrhosus*, and *C. colocynthis* are the species of the genus *Citrullus* (Chomicki and Renner 2015). Most *Citrullus* species have their origins and genetic diversity in continental Africa (Dane and Lang 2004). One theory claims that it came from a species of *Citrullus lanatus* that is a common wild plant in central Africa, while another believes that it was domesticated from the perennial *Citrullus colocynthis* that is a common plant in ancient sites. Watermelons have been grown in Africa for over 4000 years. Watermelons were commonly planted in the Nile valley region prior to 2000 BCE, according to seeds and plant pieces found in Egyptian tombs. They were transported from Africa to India in the year 800 CE and to China in the year 900 CE, after which they expanded to other continents in the year 1500s. Annual watermelon plants have lobed leaves, long, angular vines that trail, branching tendrils, and solitary male and female flowers. Watermelon of various shapes like round, oval, or elongated can weigh anywhere between 1.5 kg and 15 kg. The rind varies in color from light to dark green and has different striped patterns. Despite the fact that the flesh may be white, green, yellow, orange, or red, customers favor inner qualities like sweetness and texture and colors like deep red, pink, or dazzling yellow. Fruit types and cultivars differ widely in size and form, and the outside skin is smooth, sutured, or netted, with a white, green, or yellow tone. Normal colors for the flesh (mesocarp) are green or orange; however, pink and white are also found. Fig. 1 is a panel displaying the variety in morphology of watermelon fruits. A small genetic background of sweet watermelon has resulted from modern breeding practices that have mostly focused on fruit quality characteristics including sugar content, flesh color, and rind pattern (Levi et al. 2017). It is uncertain how phenotypic alterations brought about by human and natural choices affected the watermelon genome. For the creation and marketing of new products, sweet watermelon fruit characteristics are essential. The pharmaceutical industry, the processing industry, and the fresh market will all profit from the novel and value-added genotypes. Pickles, jam, fruit puree, popsicles, and watermelon juice are a few examples of items with added value that are made from recently created cultivars with superior qualities. The new cultivars ought to have a range of bioactive

Fig. 1 A panel showing watermelon fruit morphological diversity

substances that are both nutritive and therapeutic in nature (Mashilo et al. 2022). A viable and alternate method for hastening the creation and release of watermelon varieties with sufficient agronomic and quality traits to meet the crop's value chains is nonconventional breeding using gene-editing technology. For instance, changing watermelon genes associated with agronomic traits using clustered regularly interspaced short palindromic repeats (CRISPR/Cas9) allowed the development of novel cultivars (Wang et al. 2021). The development of superior cultivars with higher nutritional contents, market-preferred traits, and a longer shelf life will be aided by the information on the genetic control of fruit qualities in watermelons that will be possible by gene-editing and related technologies. The goal of this chapter is to illustrate the prospects and developments made in the genetic study of fruit quality characteristics in watermelon as a guide for quality breeding based on economic and end-user qualities, taking into account the aforementioned context.

2 Organic Acids and Sugar

Sugars and organic acids have a substantial influence on organoleptic fruit quality and are key components in fruit flavor development. Contrary to staple food crops, where production is the ultimate breeding goal, watermelon places greater emphasis on flavor

and aroma, both of which are influenced by the metabolite composition of the fruit. During development, watermelon fruits go through a variety of biochemical changes, including as adjustments to sugar metabolism, an increase in organic acid and color, fruit softening, flavor, and volatile aromatic compounds (Zhu et al. 2017). In one study, researchers examined the coexpression patterns of gene networks linked to sugar and organic acid metabolism using transcriptome profiles. They found three gene networks or modules comprising 2443 genes that were substantially associated with organic acids and carbohydrates. Seven more genes involved in the metabolism of organic acids and carbohydrates were found. *SWEET*, *EDR6*, and *STP* were recognized as sugar transporters (*Cla97C01G000640*, *Cla97C05G087120*, and *Cla97C01G018840*, r 2 = 0.83 with glucose content), while *Cla97C03G064990* (*Cla97C03G064990*, r 2 = 0.92 with sucrose concentration) was identified as a sucrose synthase. *Cla97C07G128420*, *Cla97C03G068240*, and *Cla97C01G008870* were identified as malate and citrate transporters, respectively, since they displayed strong associations with malic acid (r 2 = 0.75) and citric acid (r 2 = 0.85) (ALMT7, CS, and ICDH) (Umer et al. 2020a). The two enzymes that regulate sugar metabolism in watermelon most crucially are sucrose synthase and sucrose phosphate synthase (Guo et al. 2015). The expression of gene clusters involved in sugar biosynthesis, including a-galactosidase, invertase, and urease diphosphate (*UDP*)-galactose/glucose pyrophosphorylase (*UDP-Gal/Glc PPase*), rises as the watermelon fruit ripens. The *Cla013902* gene reportedly affects how sugar is metabolized in watermelons (Guo et al. 2015). There are nine a-galactosidase genes in the watermelon genome. The nine genes are reportedly involved in the hydrolysis of stachyose and raffinose, according to studies (Guo et al. 2013). Additionally, the buildup of sugar in watermelon is influenced by five genes for insoluble acid invertase (*IAI*) (Guo et al. 2015). The watermelon fruit's extracellular sucrose degeneration, which permits fructose and glucose transfer and intercellular sugar accumulation, is linked to the IAI gene *Cla020872* (Guo et al. 2015). Recent research has identified *ClAGA2*, an alkaline a-galactosidase gene expressed in the vascular bundle, as a critical regulator of the hydrolysis of stachyose and raffinose in watermelon. *ClAGA2* controls fruit raffinose hydrolysis and reduces the amount of sugar in fully grown watermelon fruits (Ren et al. 2021). Tonoplast sugar transporter (*ClTST2*) and sugar transporter 3 (*ClSWEET3*) genes control sugar storage and transfer in watermelon fruit cell vacuoles (Ren et al. 2021). Several important genes involved in sugar production and translocation that are up- or downregulated throughout developmental processes have been found by several researches. Differentially expressed genes such NAD-dependent malate dehydrogenase (*NAD-cyt MDH*), aluminum-activated malate transporter (*ALMT*), and citrate synthase (*CS*) affect the accumulation of organic acids in watermelon (Gao et al. 2018). The reversible conversion of malate to oxaloacetate is catalyzed by the NAD-dependent malate dehydrogenase (*NAD-cyt MDH*) gene (*OAA*) (Yao et al. 2011), whereas citrate synthase (*CS*) gene controls citric acid production. Malate dehydrogenase genes and aluminum-triggered malate transporters control the regulation and breakdown of malates (Umer et al. 2020a). It is believed that the genes for citrate synthase (*Cla97C03G068240*) and isocitrate dehydrogenase (*Cla97C01G008870*) are involved in the generation and breakdown, respectively, of citric acid (Umer et al. 2020b). Malic and citric acid accumulation are linked to higher

expression of the genes for malate dehydrogenases (*Cla008235* and *Cla011268*) and citrate synthases (*Cla013500*) (Gao et al. 2018). From the sweet and sweet-and-sour genotypes, *Cla97C01G000640* (*SWEET*), *Cla97C05G087120* (*EDR6*), and *Cla97C01G018840* were shown to control glucose biosynthesis, whereas *Cla97C03G064990* controls sucrose production in watermelon (Umer et al. 2020b). *ClVST1*, a gene for a vacuolar sugar transporter, was shown to be highly expressed during the ripening of watermelon fruit and was connected to the accumulation of sucrose (Ren et al. 2021).

3 Amino Acid Compositions

The amino acid citrulline is the most abundant in ripe watermelon fruit (Joshi et al. 2019). Citrulline is a nonessential amino acid which is generated throughout the urea cycle as a metabolic intermediate (Bahri et al. 2013). This amino acid serves as a precursor to arginine, another essential amino acid contained in watermelon fruit (Joshi et al. 2019). Watermelon has many genes that are involved in citrulline metabolism. Citrulline biosynthesis in watermelon is associated with ornithine carbamoyltransferase (*OTC; ClCG05G018820*), N-acetylornithine aminotransferase (*N-AOA*; *ClCG09G003180*), N-acetylornithine-glutamate acetyltransferase (*N-AOGA*), *CPS*-1 (*ClCG11G013120*) and *CPS-2* (*ClCG09G021680*), N-acetylornithine deacetylase (*AOD-1*), N-acetylornithine deacetylase (*AOD-3*), and nitric-oxide synthase gene (*ClCG01G004960*) (Joshi et al. 2019). Citrulline catabolism is also connected to ASS, 1,2,3-argininosuccinate synthase; *ASL*, *1,2-argininosuccinate lyase*; *ARG*, *arginase*; *ODC*, *ornithine decarboxylase*; and *ADC*, *arginine decarboxylase* (Joshi et al. 2019). Citrulline biosynthesis is mediated by the genes *argininosuccinate lyase*, *N-acetylglutamate kinase*, and *ornithine decarboxylase*, whereas arginine biosynthesis and accumulation is mediated by the genes *ornithine carbamoyltransferase* (Fall et al. 2019).

4 Fruit Bitterness

Watermelons in the wild create bitter cucurbitacin molecules, a kind of highly oxygenated tetracyclic triterpene that repels pests. Cucurbitacin B (CuB), cucurbitacin C (CuC), cucurbitacin E (CuE), and cucurbitacin E-2-O-glucoside are all found in *Citrullus* fruits, leaves, roots, and stems (CuE-Glu) (Kim et al. 2018). The principal bitter ingredient in *Citrullus* fruit is CuE, commonly known as elaterinide (Matsuo et al. 1999). Despite the fact that these protective chemicals evolved in plants millions of years ago, humans have bred them to make them more appealing to our taste buds. In domesticated varieties of watermelon, a single point mutation in a transcription factor results in a faulty protein and diminished bitter compounds (Everts 2016). The genetic basis of two watermelon fruit traits were studied in a backcross generation arising from the hybridization of an interspecific F, hybrid of *Citrullus lanatus* and *C. colocynthis* with the domesticated parent *Citrullus*

lanatus. Bitterness of the fruit, which distinguishes wild *C. colocynthis*, was revealed to be governed by a single dominant gene (*Bi*) that was linked to the isozyme marker *Pgm-1* at a distance of 11.3 cm (Navot et al. 1990). In one study, it was shown that *Cla011508* (located on chromosome 1) regulates the bitterness of watermelon fruit, and the crucial mutation locus in this gene provided molecular insights for marker-assisted breeding of target characteristic (Gong et al. 2022). The dominant *Bi* gene regulates cucurbitacin production, which is responsible for bitterness in *Citrullus* fruits, whereas the recessive su (bitterness suppressor) gene regulates the presence or absence of bitterness in watermelon fruit. The *Bi* gene has been identified as an oxidosqualene cyclase (*OSC; Cla007080*) gene on watermelon chromosome 6 (Chambliss et al. 1968; Robinson et al. 1976; Navot et al. 1990; Lu et al. 2016). Furthermore, one watermelon fruit bitterness gene on chromosome 1 was found; the significant locus with the highest LOD score (58.361) was designated *qbt-c1-1*, and it explained 82.927% of phenotypic variation with a negative additive effect of −0.465 (Li et al. 2018). Watermelon breeders looking to improve their carotenoid profiles should hunt for progenies with genes that condition the carotenoid synthesis pathway.

5 Fruit Shape and Size

Fruit form and size are important horticultural sector traits that exhibit a large range of phenotypic variation, emphasizing their importance in breeding programs. The shape of watermelon fruits was assumed to be controlled by an incompletely dominant gene, resulting in elongate (OO), oval (Oo), and spherical (oo) fruits (Guner and Wehner 2004). It was also shown that a single gene regulates both spherical (Os) and oval (O+) watermelon fruits, exhibiting partial dominance when a spherical fruit inbred line crosses with an oval fruit inbred line (Tanaka et al. 1995). The similar pattern of inheritance was seen in F2 populations of 'Peerless', 'Baby Delight', 'Northern Sweet', and 'Dove' (Poole and Grimball 1945). For the dominant elongate fruit, allele *ObE* was proposed; for the recessive oblong fruit, allele *Ob*; and for the round fruit, allele *ObR* (not the same as the o gene for round) (Lou and Wehner 2016). Segregation analysis in F2 and BC1 populations derived from a cross between two inbred lines 'Duan125' (elongate fruit) and 'Zhengzhouzigua' (spherical fruit) revealed that watermelon fruit shape is controlled by a single locus and that elongate fruit (OO) is only partially dominant to spherical fruit (oo), with the heterozygote (Oo) being oval fruit and a 159 bp deletion (Dou et al. 2018). The *SUN* gene, a member of the IQ domain (*IQD*) family, has long been recognized to influence tomato fruit elongation early in fruit development, following pollination and fertilization (Van Der Knaap and Tanksley 2001). Several QTLs linked with FD (and fruit weight) and FL (fruit length) have been found in diverse genetic backgrounds; however, the genes underlying these QTLs remain unknown.

6 Flesh Color

There are many different fruit flesh colors in watermelons (Fig. 1); fruit flesh color is a key feature that influences nutritional value, customer choice, and breeder selection. The composition of chlorophyll and carotenoids in the flesh determines its color. Watermelon flesh colors have been classified as scarlet red, red, pink, orange, canary yellow, pale yellow, and white based on carotenoid levels. Watermelons with red flesh (including flaming red and pink) are high in lycopene (Sun et al. 2018). Prolycopene and carotene concentrations are considerably higher in watermelons with orange flesh (Branham et al. 2017). Watermelons with yellow flesh (canary yellow and light yellow) contain a high concentration of neoxanthin, followed by neochrome and violaxanthin (Fang et al. 2020). Watermelons with white flesh have very little violaxanthin and lutein in them (Lv et al. 2015). The idea of genetic variability in the color of watermelon fruit flesh is supported by the existence of separate mechanisms regulating carotenoid metabolism. Understanding the process of carotenoid inheritance enables for the development of cultivars with improved phytochemical component compositions. The genetics of skin color are extremely complex, with numerous genes and quantitative trait loci (QTLs) influencing carotenoid production. Genes involved in the carotenoid biosynthesis and metabolism pathway's genome-wide comparative expression study showed complex gene expression and regulatory networks that led to the accumulation of different carotenoids in watermelon fruit (Fang et al. 2020; Mashilo et al. 2022). Based on genotyping data, two Kompetitive Allele-Specific PCR (KASP) markers were created for the candidate gene *Cla97C10G185970*, which was annotated as plastid lipid-associated protein and showed a strong connection between pale green and non-pale green meat fruits (Pei et al. 2021). During fruit ripening, the carotenoid profiles of four watermelon cultivars – red-fleshed 'CN66', pink-fleshed 'CN62', yellow-fleshed 'ZXG381', and white-fleshed 'ZXG507' – were examined. It was revealed that the amounts of violaxanthin and lutein in yellow fruit were positively correlated with *CHYB* and *ZEP* transcription levels (Lv et al. 2015). *ClPAPs, Cla006670, Cla010946*, *Cla008831, Cla014416, Cla021506, Cla003468*, and *Cla003198* are plastid lipid-associated genes that are thought to be involved in the development of plastoglobules and globular and crystalloid chromoplasts (Fang et al. 2020).

7 Rind Pattern

The most common rind (or skin) colors in watermelon are solid green (dark, medium, and light), striped (narrow, medium, and large dark green stripes on a light green background), and grey. Gray is also written grey, although the names of watermelons have been standardized to the grey spelling. The expression of genes responsible for rind color and pattern does not appear to be uniform across different

genetic backgrounds. Furthermore, the inheritance of gray or medium green rind colors has not been recorded, despite the fact that they have been two of the most prevalent rind colors in watermelon breeding during the previous century. ClCG08G017810 (ClCGMenG), a protein encoding a 2-phytyl-1,4-beta-naphthoquinone methyltransferase, is linked to the production of dark green rind vs light green rind in watermelon (Li et al. 2018). For the '0901', '10909', '109905', and '90509' rind trait-segregating F2 populations, genotyping analyses were done using subsets of 188, 273, 287, and 113 probes, respectively. For the '0901', '10909', '109905', and '90509' populations, 26, 34, 30, and 15 linkage groups containing 175, 254, 269, and 79 probes were created, respectively. The genetic order of the probes is mainly collinear with the physical order on the reference genome, with a few exceptions on chromosomes 1, 3, and 11. S, D, and Dgo, together with chr4 150/chr4 249 on chromosome 4 and chr6 25767 on chromosome 6, were identified nearby (Park et al. 2016). As a consequence of genetic investigations, a team of researchers found three unique genes in watermelon. In comparison to Angeleno Black Seeded, the type line for the hue of watermelon red flesh, scarlet red flesh (*Scr*) gave more vivid red color in Dixielee and Red-N-Sweet. They suggest calling the original red skin color coral red in order to distinguish it from scarlet red. As a single dominant gene, *Scr* is inherited. A single dominant gene called *Yellow Belly* (*Yb*) was identified as the cause of Black Diamond's ground spot's transition from creamy white to dark yellow. A single recessive gene called intermittent stripes (*ins*), with the dominant allele, was found to be responsible for the difference between continuous and intermittent stripes on the rind of Navajo Sweet (Gusmini and Wehner 2006).

8 Flesh Firmness

The firmness of the flesh determines the texture and quality of watermelon fruit. Multigenes control flesh firmness as a qualitative attribute. For watermelon genetic breeding, it is crucial to identify the regulatory elements that most significantly affect the firmness of the fruit's flesh. According to localization interval transcriptome analysis, *Cla012507* (MADS-box transcription factor) may be involved in the control of fruit ripening and affect the hardness of watermelon fruit. *Cla016033* (*DUF579* family member), which may impact the cell wall component contents to alter the flesh firmness in watermelon fruit, was distinct in W1-1 and PI186490 (Sun et al. 2020). The hardness of watermelon flesh is controlled by phytohormone levels, particularly ABA (Wang et al. 2017). *Cla009779* (*NCED*), *Cla005404* (*NCED*), *Cla020673* (*CYP707A*), *Cla006655* (*UGT*), and *Cla020180* (*SnRK2*) are implicated in ABA biosynthesis in watermelon (Wang et al. 2017). *Cla009779*, *Cla005404*, and *Cla005457* were the most effective at increasing ABA accumulation. On chromosome 6 of the watermelon genome here is a significant QTL (*Qffi6.1*) for flesh firmness from *C. amarus* (Gao et al. 2016). The putative candidate gene for *Qffi6.1* is *Cla018816*, a xyloglucan endotransglucosylase/hydrolase (*XTH*) gene that is variably expressed across firm- and soft-bodied near-isogenic lines (Anees et al. 2021). Another *XTH* gene

(*Cla006648*), cellulose synthase (*Cla012351*), galactosyltransferase (*Cla006648*), pectinesterase gene (*Cla004251*), ethylene response element transcription factor 1 (*Cla004120*), and ethylene response element transcription factor 2a (*Cla007092*) all played important roles in watermelon flesh firmness, according to transcriptome analysis (Anees et al. 2021).

9 Rind Thickness and Toughness

Watermelon fruit rind breaking not only facilitates disease invasion and reduces yield, but it also degrades the fruit's exterior aesthetic value. One of the simplest methods for gauging consumer acceptance during commercial purchases is the rind of a watermelon. After crossing 97103 and PI296341, a biparental F2 population was produced. Using RAPD and SSR markers, QTL analysis showed that there were a total of two rind thickness QTLs and three fruit weight QTLs (Fan et al. 2000). Additionally, a watermelon backcross (BC) generation was developed for measuring rind hardness, and a QTL location controlling rind hardness was found on chromosome 4 (Hashizume et al. 2003). Watermelon rind hardness is correlated with the ethylene *clerf4* transcription factor genes, according to the genotyping of 349-F2 individuals from 32 germplasm. The genotyping indicated a significant ascending allelic pattern of aa (hard) bb (soft) and substantial QTL region on chromosome 10 (Liao et al. 2020). Another recent study combined the hard-fleshed and soft-fleshed watermelon lines 'PI186490' and 'W1–1', and preliminary mapping in 175-F2 individuals identified the key genes on chromosomes 2 and 8 controlling central flesh hardness using BSAseq and CAPS marker-based QTL analysis (Sun et al. 2020). Using a combinatory genomic map and bulk segregant analysis, it was possible to link variations in rind hardness to the ethylene-responsive transcription factor 4 (*ClERF4*) (*BSA*). The *ClERF4* gene on chromosome 10 also has an 11-bp InDel and a neighboring SNP, which confers cracking resistance in F2 populations with different rind hardness (Liao et al. 2020). A transcriptome study demonstrated the molecular pathways involved in the enhancement of fruit attributes including greater rind toughness by watermelon and bottle gourd grafting (Garcia-Lozano et al. 2019).

10 Ethylene and Ripening in Watermelon

A microarray and quantitative real-time PCR-based investigation was conducted to better understand the sequence of events linked to fruit growth and ripening in watermelon (*Citrullus lanatus* [Thunb.] Matsum. and Nakai var. lanatus). This study found several of the ESTs with potential roles in the growth and ripening of watermelon fruits, particularly those involving the vascular system and ethylene (Wechter et al. 2008). The researchers noted differential expression of homologs of genes involved in ethylene biosynthesis (*ACC* oxidase) and signal transduction (ethylene receptor *Cm-ETR1*, ethylene insensitive [*EIN3/EIL*]-like transcription factor, ethylene-responsive binding protein [*EREBP*], and ethylene response factor [*ERF*]) in the same

study. Fruit rind plays a major role in reducing moisture loss and disease, as well as cracking resistance, ease of transport, and storage stability quality of watermelon; an ethylene-responsive transcription factor 4 (*ClERF4*) linked with variation in rind hardness was discovered using a combinatory genetic map and bulk segregant analysis (BSA) (Liao et al. 2020). The ethylene biosynthesis and signaling pathway genes, such as ACC oxidase, ethylene receptor, and ethylene-responsive factor, showed highly ripening-associated expression patterns in the watermelon, a non-climacteric fruit, according to a comparative transcriptome profiling analysis of the cultivated watermelon 97103 and wild watermelon PI296341-FR. This suggests that ethylene may play a role in the development and ripening of the fruit (Guo et al. 2015). *XTH* gene (Cla006648), cellulose synthase (*Cla012351*), galactosyltransferase (*Cla006648*), pectinesterase gene (*Cla004251*), ethylene response element transcription factor 1 (*Cla004120*), and ethylene response element transcription factor 2a (*Cla007092*) all played important roles in watermelon flesh firmness, according to transcriptome analysis (Anees et al. 2021).

11 Conclusion

Watermelon fruit qualitative characters like seed color, seed size, fruit shape, skin color, rind pattern, flesh color and sugar/acid composition, fruit bitterness, and many more exhibit significant genetic diversity. With the advancement of genomics and availability of watermelon genome sequences, it has become possible to identify genes critical to valuable fruit quality traits. As the knowledge of the molecular mechanisms behind these characteristics improves, more effective and focused selection will enhance the efficiency of breeding of this important crop.

References

Anees M, Gao L, Umer MJ, Yuan P, Zhu H, Lu X, He N, Gong C, Kaseb MO, Zhao S, Liu W (2021) Identification of key gene networks associated with cell wall components leading to flesh firmness in watermelon. Front Plant Sci 12:1164. https://doi.org/10.3389/FPLS.2021.630243/BIBTEX

Bahri S, Zerrouk N, Aussel C, Moinard C, Crenn P, Curis E, Chaumeil J-C, Cynober L, Sfar S (2013) Citrulline: from metabolism to therapeutic use. Nutrition 29:479–484

Branham S, Vexler L, Meir A, Tzuri G, Frieman Z, Levi A, Wechter WP, Tadmor Y, Gur A (2017) Genetic mapping of a major codominant QTL associated with β-carotene accumulation in watermelon. Mol Breed 37:1–13. https://doi.org/10.1007/S11032-017-0747-0

Chambliss OL, Erickson HT, Jones CM (1968) Genetic control of bitterness in watermelon fruits. In: Proceedings of the American Society for Horticultural Science. American Society for Horticultural Science, Alexandria, VA, USA, pp 539–546

Chomicki G, Renner SS (2015) Watermelon origin solved with molecular phylogenetics including Linnaean material: another example of museomics. New Phytol 205:526–532

Dane F, Lang P (2004) Comparative analysis of cpDNA variability in wild and cultivated Citrullus species: implications for evolution of watermelon. Am J Bot 91:1922–1929

Dou J, Zhao S, Lu X, He N, Zhang L, Ali A, Kuang H, Liu W (2018) Genetic mapping reveals a candidate gene (ClFS1) for fruit shape in watermelon (Citrullus lanatus L.). Theor Appl Genet 131:947–958. https://doi.org/10.1007/S00122-018-3050-5

Everts S (2016) Bitterness in wild cucumber, melon, and watermelon. Chem Eng News 94(48):10–11

Fall LA, Perkins-Veazie P, Ma G, McGregor C (2019) QTLs associated with flesh quality traits in an elite× elite watermelon population. Euphytica 215:1–14

Fan M, Xu Y, Zhang HY, Ren HZ, Kang GB, Wang YJ, Chen H (2000) Identification of quantitative trait loci associated with fruit traits in watermelon [Citrullus lanatus (Thunb) Mansf] and analysis of their genetic effects. Yi Chuan Xue Bao 27:902–910

Fang X, Liu S, Gao P, Liu H, Wang X, Luan F, Zhang Q, Dai Z (2020) Expression of ClPAP and ClPSY1 in watermelon correlates with chromoplast differentiation, carotenoid accumulation, and flesh color formation. Sci Hortic (Amsterdam) 270:109437. https://doi.org/10.1016/J.SCIENTA.2020.109437

Gao L, Zhao SJ, Lu XQ, He N, Zhu HJ, Dou JL (2016) Linkage analysis on flesh firmness of watermelon [Citrullus lanatus (Thunb.) Mansf.] by SSR molecular marker. J Plant Genet Resour 17:866–870

Gao L, Zhao S, Lu X, He N, Liu W (2018) 'SW', a new watermelon cultivar with a sweet and sour flavor. HortScience 53:895–896

Garcia-Lozano M, Dutta SK, Natarajan P, Tomason YR, Lopez C, Katam R, Levi A, Nimmakayala P, Reddy UK (2019) Transcriptome changes in reciprocal grafts involving watermelon and bottle gourd reveal molecular mechanisms involved in increase of the fruit size, rind toughness and soluble solids. Plant Mol Biol 102:213–223. https://doi.org/10.1007/S11103-019-00942-7

Gong C, Li B, Anees M, Zhu H, Zhao S, He N, Lu X, Liu W (2022) Fine-mapping reveals that the bHLH gene Cla011508 regulates the bitterness of watermelon fruit. Sci Hortic (Amsterdam) 292:110626. https://doi.org/10.1016/J.SCIENTA.2021.110626

Guner N, Wehner TC (2004) The genes of watermelon. HortScience 39:1175–1182. https://doi.org/10.21273/HORTSCI.39.6.1175

Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, Zheng Y, Mao L, Ren Y, Wang Z (2013) The draft genome of watermelon (Citrullus lanatus) and resequencing of 20 diverse accessions. Nat Genet 45:51–58

Guo S, Sun H, Zhang H, Liu J, Ren Y, Gong G, Jiao C, Zheng Y, Yang W, Fei Z, Xu Y (2015) Comparative transcriptome analysis of cultivated and wild watermelon during fruit development. PLoS One 10:e0130267. https://doi.org/10.1371/JOURNAL.PONE.0130267

Gusmini G, Wehner TC (2006) Qualitative inheritance of rind pattern and flesh color in watermelon. J Hered 97:177–185. https://doi.org/10.1093/JHERED/ESJ023

Hashizume T, Shimamoto I, Hirai M (2003) Construction of a linkage map and QTL analysis of horticultural traits for watermelon [Citrullus lanatus (THUNB.) MATSUM & NAKAI] using RAPD, RFLP and ISSR markers. Theor Appl Genet 106:779–785

Joshi V, Joshi M, Silwal D, Noonan K, Rodriguez S, Penalosa A (2019) Systematized biosynthesis and catabolism regulate citrulline accumulation in watermelon. Phytochemistry 162:129–140

Kim YC, Choi D, Zhang C, Liu H-f, Lee S (2018) Profiling cucurbitacins from diverse watermelons (Citrullus spp.). Hortic Environ Biotechnol 59:557–566. https://doi.org/10.1007/S13580-018-0066-3

Levi A, Jarret R, Kousik S, Patrick Wechter W, Nimmakayala P, Reddy UK (2017) Genetic resources of watermelon. In: Genetics and genomics of Cucurbitaceae. Springer, Cham, pp 87–110

Li B, Lu X, Dou J, Aslam A, Gao L, Zhao S, He N, Liu W (2018) Construction of a high-density genetic map and mapping of fruit traits in watermelon (Citrullus Lanatus L.) based on whole-genome resequencing. Int J Mol Sci 19:3268. https://doi.org/10.3390/IJMS19103268

Liao N, Hu Z, Li Y, Hao J, Chen S, Xue Q, Ma Y, Zhang K, Mahmoud A, Ali A, Malangisha GK, Lyu X, Yang J, Zhang M (2020) Ethylene-responsive factor 4 is associated with the desirable rind hardness trait conferring cracking resistance in fresh fruits of watermelon. Plant Biotechnol J 18:1066–1077. https://doi.org/10.1111/PBI.13276

Lou L, Wehner TC (2016) Qualitative inheritance of external fruit traits in watermelon. HortScience 51:487–496

Lu BY, Zhou HW, Chen X, Luan FS, Wang XZ, Jiang Y (2016) QTL analysis of fruit traits in watermelon. J Fruit Sci 33:1206–1218

Lv P, Li N, Liu H, Gu H, Zhao WE (2015) Changes in carotenoid profiles and in the expression pattern of the genes in carotenoid metabolisms during fruit development and ripening in four watermelon cultivars. Food Chem 174:52–59. https://doi.org/10.1016/J.FOODCHEM.2014.11.022

Mashilo J, Shimelis H, Ngwepe RM, Thungo Z (2022) Genetic analysis of fruit quality traits in sweet watermelon (Citrullus lanatus var. lanatus): a review. Front Plant Sci 13. https://doi.org/10.3389/FPLS.2022.834696

Matsuo K, DeMilo AB, Schroder RFW, Martin PAW (1999) Rapid high-performance liquid chromatography method to quantitate elaterinide in juice and reconstituted residues from a bitter mutant of Hawkesbury watermelon. J Agric Food Chem 47:2755–2759. https://doi.org/10.1021/JF9811572

Navot N, Sarfatti M, Zamir D (1990) Linkage relationships of genes affecting bitterness and flesh color in watermelon. J Hered 81:162–165. https://doi.org/10.1093/OXFORDJOURNALS.JHERED.A110952

Park S-w, Kim KT, Kang SC, Yang HB (2016) Rapid and practical molecular marker development for rind traits in watermelon. Hortic Environ Biotechnol 57:385–391. https://doi.org/10.1007/S13580-016-0005-0

Pei S, Liu Z, Wang X, Luan F, Dai Z, Yang Z, Zhang Q, Liu S (2021) Quantitative trait loci and candidate genes responsible for pale green flesh colour in watermelon (Citrullus lanatus). Plant Breed 140:349–359. https://doi.org/10.1111/PBR.12908

Poole CF, Grimball PC (1945) Interaction of sex, shape, and weight genes in watermelon. J Agric Res 71:533–552

Ren Y, Li M, Guo S, Sun H, Zhao J, Zhang J, Liu G, He H, Tian S, Yu Y (2021) Evolutionary gain of oligosaccharide hydrolysis and sugar transport enhanced carbohydrate partitioning in sweet watermelon fruits. Plant Cell 33:1554–1573

Robinson RW, Munger HM, Whitaker TW, Bohn GW (1976) Genes of the Cucurbitaceae. Hortic Sci 11:554–568

Sun T, Yuan H, Cao H, Yazdani M, Tadmor Y, Li L (2018) Carotenoid metabolism in plants: the role of plastids. Mol Plant 11:58–74. https://doi.org/10.1016/J.MOLP.2017.09.010

Sun L, Zhang Y, Cui H, Zhang L, Sha T, Wang C, Fan C, Luan F, Wang X (2020) Linkage mapping and comparative transcriptome analysis of firmness in watermelon (Citrullus lanatus). Front Plant Sci 11:831. https://doi.org/10.3389/FPLS.2020.00831/BIBTEX

Tanaka T, Wimol S, Mizutani T (1995) Inheritance of fruit shape and seed size of watermelon. J Jpn Soc Hortic Sci 64:543–548. https://doi.org/10.2503/JJSHS.64.543

Umer MJ, Bin Safdar L, Gebremeskel H, Zhao S, Yuan P, Zhu H, Kaseb MO, Anees M, Lu X, He N, Gong C, Liu W (2020a) Identification of key gene networks controlling organic acid and sugar metabolism during watermelon fruit development by integrating metabolic phenotypes and gene expression profiles. Hortic Res 7. https://doi.org/10.1038/S41438-020-00416-8/41991606/41438_2020_ARTICLE_416.PDF

Umer MJ, Gao L, Gebremeskel H, Bin SL, Yuan P, Zhao S, Xuqiang L, Nan H, Hongju Z, Liu W (2020b) Expression pattern of sugars and organic acids regulatory genes during watermelon fruit development. Sci Hortic (Amsterdam) 265:109102. https://doi.org/10.1016/j.scienta.2019.109102

Van Der Knaap E, Tanksley SD (2001) Identification and characterization of a novel locus controlling early fruit development in tomato. Theor Appl Genet 103:353–358. https://doi.org/10.1007/S001220100623

Wang Y, Guo S, Tian S, Zhang J, Ren Y, Sun H, Gong G, Zhang H, Xu Y (2017) Abscisic acid pathway involved in the regulation of watermelon fruit ripening and quality trait evolution. PLoS One 12:e0179944. https://doi.org/10.1371/JOURNAL.PONE.0179944

Wang Y, Wang J, Guo S, Tian S, Zhang J, Ren Y, Li M, Gong G, Zhang H, Xu Y (2021) CRISPR/Cas9-mediated mutagenesis of ClBG1 decreased seed size and promoted seed germination in watermelon. Hortic Res 8. https://doi.org/10.1038/S41438-021-00506-1/42043543/41438_2021_ARTICLE_506.PDF

Wechter WP, Levi A, Harris KR, Davis AR, Fei Z, Katzir N, Giovannoni JJ, Salman-Minkov A, Hernandez A, Thimmapuram J, Tadmor Y, Portnoy V, Trebitsh T (2008) Gene expression in developing watermelon fruit. BMC Genomics 9:1–13. https://doi.org/10.1186/1471-2164-9-275/FIGURES/5

Yao Y-X, Li M, Zhai H, You C-X, Hao Y-J (2011) Isolation and characterization of an apple cytosolic malate dehydrogenase gene reveal its function in malate synthesis. J Plant Physiol 168: 474–480

Zhu Q, Gao P, Liu S, Zhu Z, Amanullah S, Davis AR, Luan F (2017) Comparative transcriptome analysis of two contrasting watermelon genotypes during fruit development and ripening. BMC Genomics 18:1–20

Grapes: A Crop with High Nutraceuticals Genetic Diversity

Javier Tello, Loredana Moffa, Yolanda Ferradás, Marica Gasparro, Walter Chitarra, Rosa Anna Milella, Luca Nerva, and Stefania Savoi

Contents

Javier Tello and Loredana Moffa contributed equally with all other contributors.

J. Tello
Instituto de Ciencias de la Vid y del Vino (ICVV), Consejo Superior de Investigaciones Científicas – Universidad de la Rioja – Gobierno de La Rioja, Logroño, Spain
e-mail: javier.tello@icvv.es

L. Moffa
Council for Agricultural Research and Economics (CREA) – Research Centre for Viticulture and Enology, Conegliano, Italy
e-mail: loredana.moffa@crea.gov.it

Y. Ferradás
Instituto de Ciencias de la Vid y del Vino (ICVV), Consejo Superior de Investigaciones Científicas – Universidad de la Rioja – Gobierno de La Rioja, Logroño, Spain

Faculty of Biology, University of Santiago de Compostela, Santiago de Compostela, Spain
e-mail: yolanda.ferradas@icvv.es

M. Gasparro · R. A. Milella
Council for Agricultural Research and Economics (CREA) – Research Centre for Viticulture and Enology, Turi, Italy
e-mail: marica.gasparro@crea.gov.it; rosaanna.milella@crea.gov.it

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_37

Abstract

Grapevine is considered the most important fruit crop cultivated in temperate regions and is acknowledged as a model species for non-climacteric fleshy fruits. From a nutritional perspective, grapes are fruits with a high content of carbohydrates, a good nutritional source of minerals and vitamins, and most importantly, they are one of the richest fruits in polyphenols and other compounds with antioxidant properties. In particular, this chapter focuses on nutraceutical compounds such as phenolic acids, stilbenes, flavonols, flavanols, tannins, anthocyanidins, monoterpenes, sesquiterpenes, carotenoids, C_{13}-norisoprenoids, and some vitamins. Their chemical structures and biosynthetic pathways are revised, and the content and diversity of these secondary metabolites in genetic resources of the genus *Vitis* (including *Muscadinia* and *Vitis* species, focusing on *V. vinifera* cultivars) are shown. In addition, QTL and association studies exploring the genetic basis of the biosynthesis of different health-related compounds in *V. vinifera* grape berries were asserted. Finally, a survey on the peculiarities and limits of traditional breeding compared to the innovative plant breeding techniques (cisgenesis and genome editing) applied to grapevine is provided. We

W. Chitarra · L. Nerva (✉)
Council for Agricultural Research and Economics (CREA) – Research Centre for Viticulture and Enology, Conegliano, Italy

Institute for Sustainable Plant Protection, CNR, Torino, Italy
e-mail: walter.chitarra@crea.gov.it; luca.nerva@crea.gov.it

S. Savoi (✉)
Department of Agricultural, Forest and Food Sciences, University of Turin, Grugliasco, Italy
e-mail: stefania.savoi@unito.it

conclude with the potential role of grape miRNA on human health as likely candidates for dietary therapy approaches due to their cross-kingdom abilities and regulation activity of gene expression and cellular processes in humans through dietary intake.

Keywords

Vitis vinifera · Stilbenes · Anthocyanins · Monoterpenes · Carotenoids · miRNA · Cisgenesis · Genome editing

1 Introduction to Worldwide Wine and Table Grape Production

Grapevines are one of the most valuable horticulture crops, typically cultivated in regions with mild climate conditions, sufficient heat accumulation, and moderate winter low temperatures for proper fruit development and growth. The primary grape production areas are located between latitudes 30°–50° and 30°–40° in the Northern and Southern Hemispheres, respectively. These areas include some world-renowned winemaking regions, like Bordeaux and Burgundy in France, La Rioja in Spain, Tuscany in Italy, the Napa Valley in the USA, the Barossa Valley in Australia, or the Stellenbosch region in South Africa, to cite a few. It also embraces major table grape and raisin producers, including China, Turkey, India, the USA, and Chile (OIV 2019). In 2020, the total vineyard surface for wine grape and table grape production was estimated at 7.3 million hectares (OIV 2019). Mediterranean Sea countries, where grapevines have been cultivated for centuries, are some of the most important worldwide leading grape growers, including Spain (13.1% of the vineyard surface in 2020), France (10.9%), Italy (9.8%) and Turkey (5.9%). Outside these traditional grape-growing regions, it is worth highlighting other major grape-producing countries like China (10.7%), the USA (5.5%), Argentina (2.9%), and Chile (2.8%). These "New World" regions accounted for some of the largest vineyard surface and gross grape production increases over the last decades (OIV 2019).

Worldwide grape production is estimated at 77–78 million tons per year. About 57% of this annual yield mainly aims to sustain worldwide wine production. In contrast, fresh grape and dried grape (raisins) markets account for 36% and 7% of the total annual grape production, respectively (OIV 2019). In addition, some of these grapes are processed into jam, grape juice, vinegar, jelly, and grape seed extracts and oils, which diversify the use of grapes. Although some muscadine grapes (from *Vitis rotundifolia* cultivars and interspecific hybrids) have local relevance in the Southeastern USA (Yuzuak and Xie 2022), worldwide grape production is majorly sustained by the cultivation of *V. vinifera* cultivars. Despite the vast number of cultivars available for this species (Wolkovich et al. 2018), grape production focuses on cultivating a few genotypes (Morales-Castilla et al. 2020). In this light, only 13 grape cultivars ('Kyoho', 'Cabernet Sauvignon', 'Sultanina', 'Merlot', 'Tempranillo', 'Airén', 'Chardonnay', 'Syrah', 'Red Globe', 'Garnacha Tinta', 'Sauvignon Blanc', 'Pinot Noir',

Fig. 1 Grape clusters at maturity of 'Kyoho' (left), 'Carbernet Sauvignon' (middle), and 'Sultanina' (right). (Photos were downloaded from the *Vitis* International Variety Catalogue (*V*IVC), www.vivc.de – (accessed April 2022). Source: Ursula Brühl, Julius Kühn-Institut (JKI), Federal Research Centre for Cultivated Plants, Institute for Grapevine Breeding Geilweilerhof – 76,833 Siebeldingen, Germany)

and 'Trebbiano Toscano' (syn. 'Ugni Blanc')) account for more than one-third of global vineyard surface, and 33 cultivars for one half of it (Alston and Sambucci 2019). Attending to this list, the most widely cultivated genotype is 'Kyoho' (a black-berried table grape cultivar bred by Yasushi Ohinoue in 1935 in the Oinoue Institute for Agronomical and Biological Science, Japan), followed by 'Cabernet Sauvignon' (a black-berried wine grape cultivar), and 'Sultanina' (a seedless white-berried multi-purpose cultivar also known as 'Thompson Seedless') (Fig. 1). Collectively, these three cultivars accounted for almost one million hectares in 2015 (Alston and Sambucci 2019).

Global table grape production has increased gradually in the last decades to reach ca. 27 million tons in 2018, with four countries producing more than 50% of world production: China (9.5 million tons), Turkey (1.9), India (1.9), and Iran (1.7) (OIV 2019). Besides dominating table grape production, these countries also lead table grape consumption rates, likely due to the perishable nature of grape fruits, which tend to be consumed close to where they are grown. World raisins production and consumption have stayed relatively constant in the last decades, reaching a total production of 1.3 million tons in 2018 (OIV 2019). The two principal raisin producers worldwide are Turkey and the USA, with 381 and 263 thousand tons of gross production in 2018, respectively (OIV 2019). In contrast to table grapes, raisins can be easily transported, so they are majorly exported to other countries. In fact, although European countries are only minor producers of raisins, they account for one-third of global consumption (FAO and OIV 2016). As a result, raisins are considered valuable agricultural commodities for producing countries. Table grape production is based on cultivating a series of high-yielding cultivars with good sensory attributes and optimum commercial characteristics for packaging and

transport. On the other hand, raisin production is based on growing cultivars with berries with a good attitude towards dehydration and easy detachment from the stalk. For both uses, seedless cultivars are preferred. Some of the most cultivated cultivars for table grapes and raisins production are 'Alphonse Lavallée', 'Crimson Seedless', 'Dattier de Beyrouth' (syn. 'Afus Ali'), 'Flame Seedless', 'Kyoho', 'Muscat Hamburg', 'Italia', 'Muscat of Alexandria', 'Red Globe', 'Sugraone', 'Sultanina', and 'Victoria' (FAO and OIV 2016). As observed, modern table grape and raisins production is based on the combined cultivation of traditional cultivars (e.g., 'Sultanina', 'Muscat of Alexandria') with others obtained in more recent breeding programs aimed to achieve novel grapes with better features that fit consumers and producers' needs (e.g., 'Kyoho', 'Red Globe').

2 Grape as a Source of Nutraceutical Compounds

Grapes and raisins are grown on all inhabited continents, so they are commonly included in worldwide diets. Therefore, grapes result as one of the most regularly consumed fruits worldwide. It has been estimated that the average world consumption of grapes per year is 4.0 kg per capita (FAO and OIV 2016), although some differences between regions exist. For example, Turkey and China's annual grape consumption is much higher, estimated at 23 and 7 kg per capita, respectively (FAO and OIV 2016). Grapes are known to be one of the fruits with higher content of carbohydrates, mostly simple sugars (18.1 g per 100 g of grapes). As observed in Table 1, grapes are a good dietary source of calcium, manganese, phosphorous,

Table 1 Nutritional facts of grapes, raisins, and other relevant fruit crops. Nutritional data is referred to as 100 g servings. Information is taken from the USDA-Agricultural Research Service, FoodData Central (https://fdc.nal.usda.gov/, access: April 2022)

	Apple	Banana	Cherry	Grape	Pear	Raisin
Energy (kcal)	52	89	63	69	57	296
Carbohydrates (g)	13.8	22.8	16	18.1	15.2	78.5
Proteins (g)	0.26	1.09	1.06	0.7	0.36	2.52
Lipids (g)	0.17	0.33	0.2	0.2	0.14	0.54
Fiber (total, g)	2.4	2.6	2.1	0.9	3.1	6.8
Calcium (% DRV)[a]	0.6	0.5	1.4	1.1	0.9	2.9
Magnesium (% DRV) [1]	1.4	7.7	3.1	2.0	2.0	8.6
Manganese (% DRV) [1]	1.2	9.0	2.3	2.4	1.6	8.9
Phosphorous (% DRV) [1]	2.0	4.0	3.8	3.6	2.2	13.6
Potassium (% DRV) [1]	3.1	10.2	6.3	5.5	3.3	23.6
Vitamin B1 (% DRV) [1]	1.8	2.7	2.7	6.4	0.9	10.0
Vitamin B6 (% DRV) [1]	2.4	21.8	2.9	5.3	1.8	11.2
Vitamin C (% DRV) [1]	4.2	7.9	6.4	2.9	3.9	4.9
Vitamin E (% DRV) [1]	1.4	0.8	0.5	1.5	0.0	*n.a.*

[a] Dietary Reference Values (DRVs) were taken from EFSA, considering a 40-year-old man. *n.a.*: not available

potassium, and vitamins B1, B6, and E compared to other relevant fruit crops. However, grapes stand out as one of the fruits richest in polyphenols and other compounds with antioxidant properties. In this line, multiple bioactive compounds with antioxidant potential have been discovered in grapes, including, but not limited to, phenolic acids, stilbenes, anthocyanins, flavonols, and flavanols (Teixeira et al. 2013; Pinasseau et al. 2017).

Nowadays, there is a growing interest in consuming food products naturally rich in antioxidants (like grapes), as they may reduce the incidence of some chronic diseases. The interest in uncovering the connection between grape consumption and its beneficial effects on human health started in the early 1990s when the so-called "French Paradox" was proposed. This concept refers to the unusually low rate of coronary heart disease observed in French people, regardless of consuming a diet rich in high saturated fats, which was initially correlated with moderate red wine intake (Renaud and de Lorgeril 1992). Later on, this phenomenon was linked to the presence of resveratrol in the skin of colored grapes and then to multiple dietary factors and life conditions that characterize Mediterranean and French lifestyles (Ndlovu et al. 2019). These findings boosted numerous research studies aimed at identifying and quantifying the most important health-related metabolites present in grapes and their derived products and analyzing their consumption's effect on human health (Catalgol et al. 2012). Numerous epidemiological, in vivo, and in vitro studies have indicated a relationship between polyphenol grape consumption and the lower incidence of some diseases and health disorders, information that has been collected in numerous reviews (Catalgol et al. 2012; Yang and Xiao 2013; Wightman and Heuberger 2014; Singh et al. 2015; Rasines-Perea and Teissedre 2017; Herman et al. 2018; Nash et al. 2018; Ko and Kim 2020). On the contrary, very few in vivo studies have considered the effects of fresh table grape intake on human health. As an example, Ammollo and colleagues reported an anticoagulant and profibrinolytic effect in healthy subjects after a 3-week table grape-rich diet (Ammollo et al. 2017). This protective role was further linked to an effect on the modulation of a series of genes implicated in critical processes like immune response, DNA and protein repair, autophagy, and mitochondrial biogenesis (Milella et al. 2020). Another study has suggested a link between the intake of fresh table grapes and a series of microRNA involved in fighting cancer development (Tutino et al. 2021).

Grape polyphenols have been linked to a lower risk of cardiovascular diseases, one of most industrialized countries' dominant causes of death. Among other cardioprotective effects, grape flavonoids (like anthocyanins, flavanols, and flavonols) are known to (i) inhibit platelet aggregation, (ii) decrease blood levels of triglycerides and high-density lipoprotein cholesterol, (iii) increase blood levels of low-density lipoprotein cholesterol, (iv) reduce oxidative stress, and (v) improve endothelial function (Wightman and Heuberger 2014). Besides, grape consumption may have a beneficial effect on type-2 diabetes, as they have one of the lowest glycemic index and glycemic load values among fruit crops (FAO and OIV 2016), and they are rich in phenolic compounds with some capability to reduce blood glycemic levels after food intake (Yang and Xiao 2013). On the other hand, grape berries are rich in resveratrol. Besides its protective effect against cardiovascular

diseases, resveratrol has been suggested to have chemopreventive and chemotherapeutic effects on different cancer types (Catalgol et al. 2012). In addition, the estrogenic activity of resveratrol has been associated with the prevention of some metabolic disorders in post-menopausal women, including the loss of bone mineral density and some alterations of the lipid metabolism pathways that might lead to the development of insulin resistance, abdominal adiposity, and dyslipidemia (Ko and Kim 2020). Lastly, grape polyphenols have been associated with some preventive effects toward developing neurodegenerative disorders, like Alzheimer's and Parkinson's diseases, and chronic traumatic encephalopathy (Herman et al. 2018). Delphinidin 3-*O*-glucoside is one of the predominant plant bioactive compounds of anthocyanins, linked to protecting against thrombosis and cardiovascular diseases, inhibiting platelet activation, and attenuating arterial and venous thrombus formation and development (Yang et al. 2012). Flavonoids can scavenge free radicals; therefore, they play a role in protecting against UV light damage. These compounds, accumulated preferentially in the skin and seeds of ripened grape berries can fulfill these functions thanks to the flavonoid scaffolds modification, which ameliorates the antioxidant capacity among other properties such as stability, solubility, and bioavailability of the resulting derivative.

Grape seeds are especially rich in proanthocyanidins, also known as condensed tannins or tannins. Given their redox potential and capability to bind target proteins and regulate cell signaling pathways, proanthocyanidins are known to have beneficial effects against different diseases (Unusan 2020). For example, in vivo and in vitro experiments have indicated that seed proanthocyanidin can inhibit metastatic processes, positively affecting the progression of lung, breast, colon, prostate, liver, pancreas, and skin cancers (Unusan 2020). In fact, proanthocyanidins have been suggested as powerful candidate drugs for cancer prevention and treatment therapy. In addition, grape seed proanthocyanidins have shown some preventive effects against cardiovascular diseases, obesity, type-2 diabetes, inflammatory bowel disease, neurodegenerative disorders, asthma, eye diseases, and osteoarthritis (Unusan 2020). Because of these potential beneficial effects on human health, grape seed oils rich in proanthocyanidins and other antioxidants have been recommended as suitable dietary supplements to prevent certain diseases (Gupta et al. 2020). Nevertheless, detailed safety analyses are needed to set the conditions of use and consumption of grape-derived ingredients in humans, especially for subjects under specific medical conditions (Singh et al. 2015).

Lastly, grapevine leaves have a long history as a traditional food in some regions, being considered a healthy food item with multiple nutritional benefits. In addition, they are often mentioned in folk medicine to treat some medical conditions, like high blood pressure, diabetes, and several circulatory system and inflammatory disorders (Pintać et al. 2019). Analyses of their chemical composition have indicated that they are rich in carotenoids (like β-carotene, lutein, and zeaxanthin) and essential minerals like calcium, potassium, iron, and zinc (Maia et al. 2021). In addition, they show a high content of flavonoids and hydroxycinnamic acid derivatives, which confer antioxidant protective activities (Pintać et al. 2019; Banjanin et al. 2020). In this regard, preliminary in vitro studies suggested that grapevine leaf extracts might

exert some promising anti-proliferative effects in human cancer cell lines due to their phenolic composition (Ferhi et al. 2019). However, further research is needed to suggest a chemopreventive and/or chemotherapeutic effect of grapevine leaf extracts, as well as before their use as ingredients in functional foods.

Biomarkers of food intake provide accurate information on the consumption of certain foods in the diet. Indeed, several protocols for identifying and confirming food candidate biomarkers have been established within the Food Biomarker Alliance (FoodBAll) Project funded by the Initiative "A Healthy Diet for a Healthy Life" (JPI-HDHL). It was developed to cover the food intake in different European populations to identify more straightforward sampling techniques for metabolomics analysis and discover dietary biomarkers. In particular, resveratrol and tartaric acid were found as candidate biomarkers for grapes intake. However, considering the putative biomarkers for berries in general (i.e., including strawberry, blueberry, blackberry, cranberry, raspberry, and blackcurrant), it was possible to identify that anthocyanins and ellagitannins are the main drivers in berries (Ulaszewska et al. 2020).

3 Polyphenolic Compounds in Grapes

Phenolic compounds significantly impact fruit and wine quality. Their biosynthetic pathways are tightly controlled by many transcription factors that modulate their expression and accumulation as a natural defense to various biotic and abiotic stresses (Kobayashi et al. 2004; Bogs et al. 2007; Terrier et al. 2009; Czemmel et al. 2009; Höll et al. 2013). In grapevine, phenolic compounds are conventionally separated into two classes: (i) non-flavonoid (phenolic acids and stilbenes) and (ii) flavonoid compounds (flavonols, flavanols, and anthocyanins).

3.1 Phenolic Acids

Polyphenol biosynthesis begins with phenylalanine. This amino acid can exert carbon competition phenomena between primary and secondary metabolites linking the shikimate pathway with the non-oxidative branch of the pentose phosphate pathway. The biosynthesis is divided into the phenylpropanoid and flavonoid/stilbene pathways. The phenylpropanoid ends with *p*-coumaryl-CoA in three reactions carried out by the phenylalanine ammonia-lyase, the cinnamate4-hydroxylase, and the 4-coumarate-CoA ligase activities (*PAL*, *C4H*, *4CL*, respectively). *p*-coumaryl-CoA is a branching compound of the phenylpropanoid pathway, and it can set up the flavonoid (with the enzymes chalcone synthase, *CHS*) or stilbene (with the enzyme stilbene synthase, *STS*) biosynthesis pathways (Teixeira et al. 2013; Falchi et al. 2019).

Phenolic acids (coumaric, caffeic, and ferulic) can be synthesized via phenylpropanoid products. When esterified with tartaric acid, they are known as *trans*-caftaric, *trans*-coutaric, and *trans*-fertaric acids, respectively (Fig. 2). They are synthesized in the berry flesh before véraison (beginning of grape ripening). Another significant phenolic acid in grapes is gallic acid (Fig. 2). It is formed from an

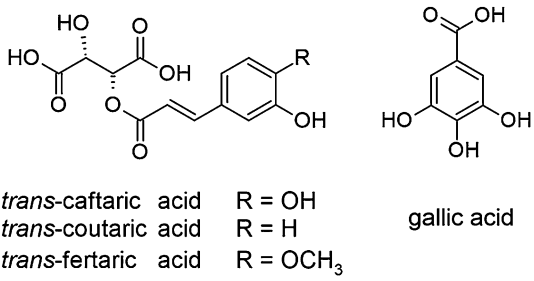


Fig. 2 Chemical structures of the primary phenolic acids in grapes. (Drawn by the authors with ChemSketch (ACD/Labs))

intermediate compound in the upstream reactions of the shikimate pathway, thanks to the activity of a dehydroshikimate dehydrogenase (Bontpart et al. 2016). Gallic acid participates in several reactions, such as the galloylation of flavanols.

3.2 Stilbenes

Stilbenes are phytoalexins, small lipophilic compounds with well-known benefits for human health. They are naturally formed by the stilbene synthases in berries from *véraison* onwards. Stilbenes content and composition in grape berries present differences among cultivars. In addition, an enhancement of their synthesis happens concurrently with biotic and abiotic stresses (Gatto et al. 2008). Genomics analysis in *V. vinifera* indicates a remarkable expansion of the stilbene synthase gene family compared to other crops and plant model species, likely through segmental and tandem gene duplication events. In fact, up to 45 stilbene synthases have been described, with 33 coding for functional proteins (Vannozzi et al. 2012). The major stilbenes in *Vitis* are *trans*-resveratrol and its isomer *cis*-resveratrol, although the latter is less stable. Modifications of resveratrol give rise to other stilbenes present in grapes, such as the glycosylated *trans*-piceid, the methoxylated pterostilbene, or the hydroxylated piceatannol, which, if further glycosylated, become astringin. Moreover, oligomerization processes can produce the resveratrol dimer pallidol or a series of viniferins (α-, β-, γ-, δ-, ε-) with different stilbene units (Fig. 3). The genetic and metabolic mechanisms involved in such modifications are mainly unknown, except for a resveratrol *O*-methyltransferase and a resveratrol glucosyl transferase responsible for the production of pterostilbene and piceid, respectively (Hall and De Luca 2007; Schmidlin et al. 2008).

3.3 Flavonols

The flavonoid pathway leads to three major polyphenolic classes: flavonols, flavanols, and anthocyanins. First, naringenin is produced through the activity of chalcone synthases and chalcone isomerases (*CHS*, *CHI*). Then, naringenin is hydroxylated at diverse positions (3, 3′ or 3′5′) thanks to different flavanone

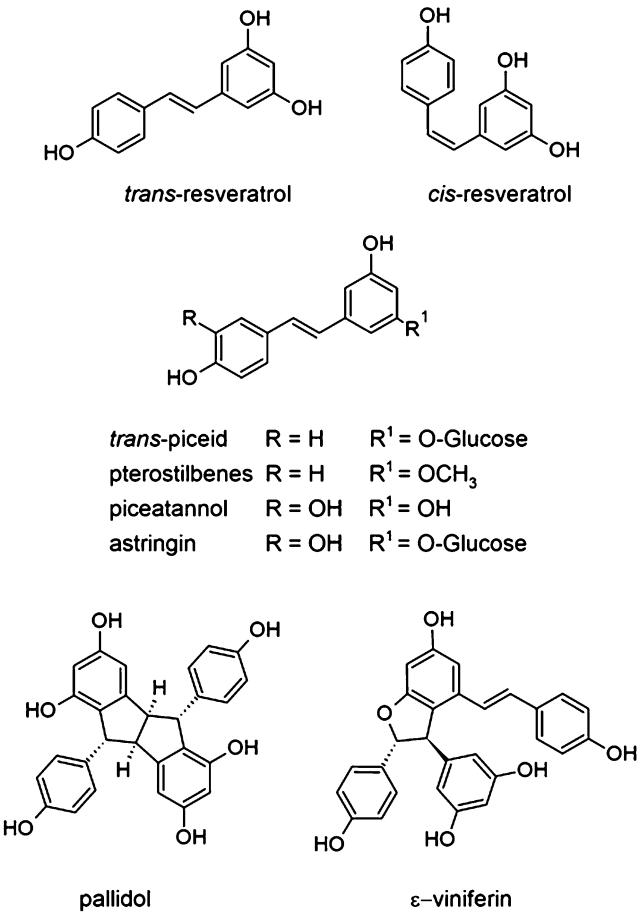
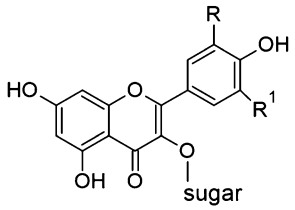


Fig. 3 Chemical structures of the major stilbenes in grapes. (Drawn by the authors with ChemSketch (ACD/Labs))

hydroxylases (*F3H*, *F3′H*, *F3′5′H*). These reactions form dihydroquercetin, dihydrokämpferol, or dihydromyricetin, respectively. Flavonol synthases (*FLS*) trigger flavonols' production by converting the previous dihydro-forms into kämpferol, quercetin, and myricetin. In addition to these three major compounds, grapes have isorhamnetin (the quercetin *O*-methylated form), laricitrin, and syringetin (the myricetin *O*-methylated forms carrying one or two methyl groups, respectively). Glycosylated flavonols are the majority metabolites found in grapes, and they can be linked to a glucoside, a galactoside, a rhamnoside, a rutinoside, or a glucuronide. In black-berried cultivars, flavonols can associate with anthocyanins, strengthening red wine color stability. In white-berried grapes, the primary flavonols are kämpferol (mono-hydroxylated), quercetin, and isorhamnetin (di-hydroxylated). In addition, myricetin, laricitrin, and syringetin (tri-hydroxylated) have been



quercetin	R = OH	R^1 = H
kämpferol	R = H	R^1 = H
myricetin	R = OH	R^1 = OH
isorhamnetin	R = OCH_3	R^1 = H
laricitrin	R = OCH_3	R^1 = OH
syringetin	R = OCH_3	R^1 = OCH_3

Fig. 4 Chemical structures of the noteworthy flavanols in grapes. (Drawn by the authors with ChemSketch (ACD/Labs))

detected in black-berried grapes (Mattivi et al. 2006) (Fig. 4). Synthetized prevalently in grape berries skins, they are suggested to have a protective effect against UV radiation.

3.4 Flavanols

Flavanols, synthesized during the first growth phase of the berry, are derived from dihydroflavonols in two ways. They can be formed (i) in three steps via the dihydroflavonol 4-reductase, the leucoanthocyanidin dioxygenases, and the anthocyanidin reductase passing by anthocyanidins (*DFR*, *LDOX*, *ANR*) or (ii) in two steps via the dihydroflavonol 4-reductase and leucoanthocyanidin reductase (*DFR*, *LAR*) (Falchi et al. 2019).

The most abundant monomeric flavanols found in *Vitis* are catechin and gallocatechin, their enantiomers epicatechin and epigallocatechin, and the gallate ester epicatechin-3-*O*-gallate (Fig. 5). Their difference lies in their chemical modification (hydroxylation, acylation) and stereochemistry. Proanthocyanidins are the major polymeric phenolics, which differ in the number and selection of monomeric flavanols (polymerization degree) and the galloylation level (Mattivi et al. 2009). Tannins can be hydrolyzed under favorable pH conditions in gallic acid (from the gallotannins) or in ellagic acid (from the ellagitannins) (Teixeira et al. 2013).

3.5 Anthocyanins

Anthocyanins are a series of pigmented metabolites present in black-berried grapes, which are synthesized in the skin and stored in the vacuole during the second phase of berry growth. The UDP-glucose:flavonoid-3-*O*-glucosyltransferase (*UFGT*) is in

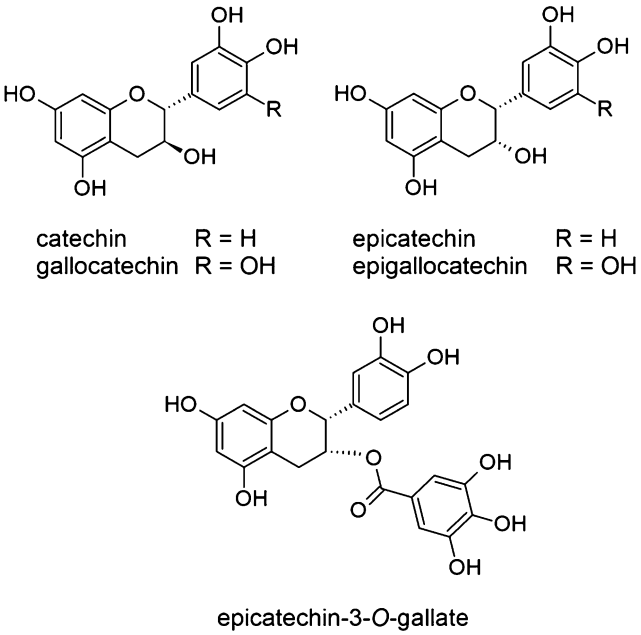
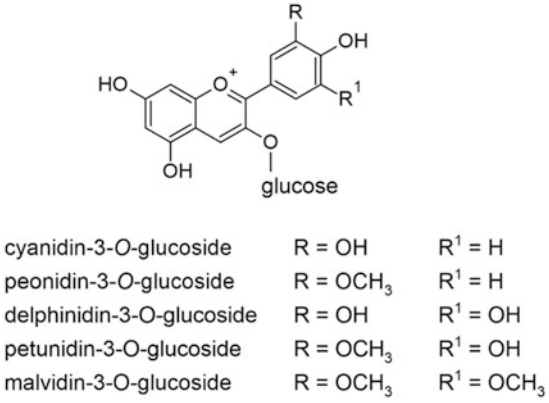


Fig. 5 Chemical structures of the main flavanols in grapes. (Drawn by the authors with ChemSketch (ACD/Labs))

Fig. 6 Chemical structures of the most important glucosylated anthocyanins in grapes. (Drawn by the authors with ChemSketch (ACD/Labs))



charge of their stabilization by glycosylation. The main grapevine anthocyanins are the cyanidin-3-*O*-glucoside and peonidin-3-*O*-glucoside (di-hydroxylated forms), and the delphinidin-3-*O*-glucoside, petunidin-3-*O*-glucoside, and malvidin-3-*O*-glucoside (tri-hydroxylated forms). In addition, peonidin, petunidin, and malvidin can further be methylated, increasing their stability and promptly influencing the berry color, (Fig. 6).

4 Terpenoid Compounds in Grapes

Terpenoid compounds can be obtained from two unrelated pathways: the plastid MEP (methyl-erythritol-4-phosphate) pathway or the cytosolic MVA (mevalonate) pathway. In any case, both pathways end with the synthesis of isopentenyl pyrophosphate (C5), which represents the building block of all terpenes, whose synthesis happens via specific prenyltransferases. Monoterpenes (C10) and tetraterpenes (C40) are produced in the plastid by geranyl pyrophosphate synthase and geranylgeranyl pyrophosphate synthase, while sesquiterpenes (C15) are formed in the cytosol by the farnesyl pyrophosphate synthase.

4.1 Monoterpenes

Up to 60 putatively terpene synthases (*TPS*) have been identified in the *vinifera* genome, some of them already functionally characterized (Martin et al. 2010). For more information, readers are referred to a recent review by Lin et al. (2019). Monoterpenes (C10) are compounds recognized as being involved in grape aroma, providing flower and fruity notes. The most aromatic monoterpenes in grapes are linalool (and its oxides, furanoid and pyranoid), geraniol (and its isomer nerol), citronellol, ho-trienol, α-terpineol and terpinen-4-ol, and *cis*-rose oxide (Mateo and Jiménez 2000) (Fig. 7). Monoterpenes diversity in grape berries arises from the linalool oxidative metabolism (Ilc et al. 2016), although the enzymes involved in this process have not yet been identified. In fact, geraniol can be converted to nerol or reduced to citronellol by an unknown isomerase and reductase, respectively. Moreover, hydroxylation and cyclization of citronellol can form *cis* and *trans*-rose oxide (Luan et al. 2005). The majority of monoterpenes are glycosylated in grapevine. Up to now, only three genes encoding for glycosyltransferases have been indicated to participate in this process (Bönisch et al. 2014a,b; Li et al. 2017). In their glycosylated forms, molecules are bound to sugar moiety and are not perceived, therefore, named as the nonvolatile fraction. Several hydrolysis mechanisms (acid or enzymatic) can free these bound terpenes during must fermentation or wine aging.

4.2 Sesquiterpenes

Sesquiterpenes (C15) are suggested to play a less important role in grape aroma, as they are generally found in concentrations lower than the olfactory perception threshold. Rotundone has been highly investigated in recent years (Fig. 7) due to the peppery character it confers to some black-berried (e.g., 'Cabernet Sauvignon', 'Syrah') and white-berried (e.g., 'Grüner Veltliner', 'Riesling) grape cultivars, and derived wines (Siebert et al. 2008; Caputi et al. 2011). Interestingly, the rotundone biosynthesis has been indicated to be linked to a specific terpene synthase coding for α-guaiene, the rotundone precursor (Drew et al. 2015).

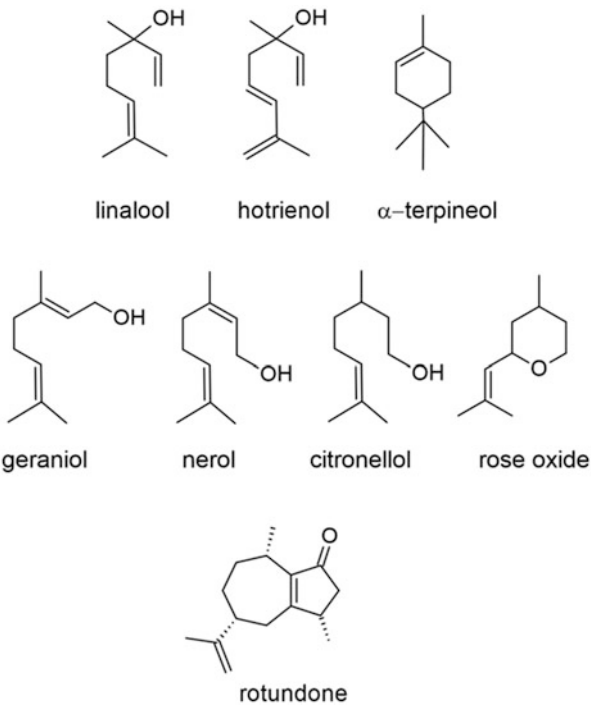


Fig. 7 Chemical structures of the major monoterpenes and sesquiterpenes in grapes. (Drawn by the authors with ChemSketch (ACD/Labs))

4.3 Tetraterpenes

Tetraterpenes in grapes are represented by carotenoids (C40), a group of pigments commonly found in different flowers and fruits. They participate in light absorption processes and, due to their antioxidant properties, shield the photosystems from photooxidative degradation. The grapevine carotenoid metabolic pathway has been characterized by a comparative genomics approach (Young et al. 2012). Following this work, phytoene is formed by joining two molecules of geranylgeranyl pyrophosphate by phytoene synthase (*PSY*). In this regard, this first step is suggested to be the rate-limiting reaction of carotenoid biosynthesis. Then, phytoene is modified by four consecutive desaturation reactions by a phytoene desaturase, a carotene isomerase, a carotene desaturase, and a carotenoid isomerase (*PDS*/*PDH*, *ZISO*, *ZDS*, *CISO*), that end with the production of lycopene. Then, lycopene cyclases (*LBCY*) form α- and β-carotene, which triggers the formation of a series of carotenoid compounds, such as lutein and zeaxanthin, antheraxanthin and violaxanthin through different hydroxylation reactions (Fig. 8). The conversion of violaxanthin into neoxanthin represents the last reaction in the core carotenoid biosynthetic pathway, which is done by a neoxanthin synthase (*NSY*). The process of metabolite

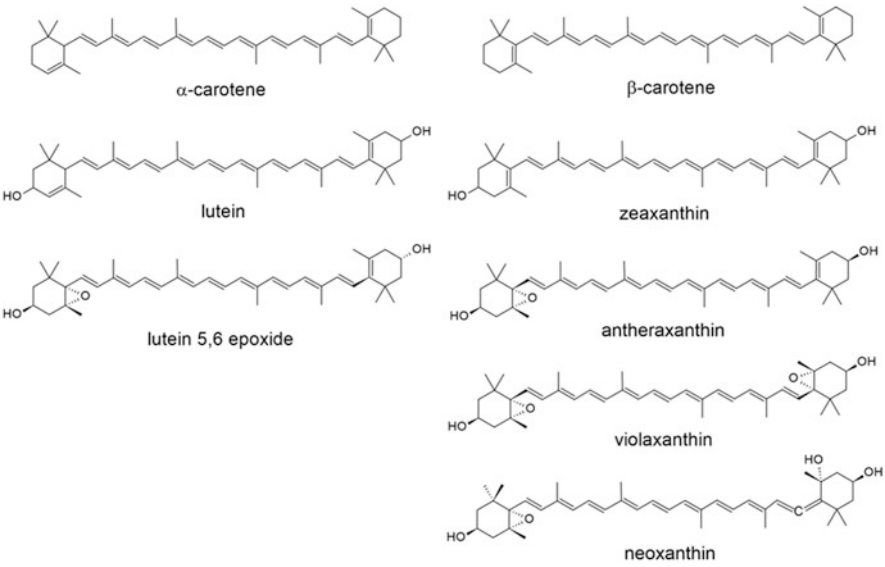
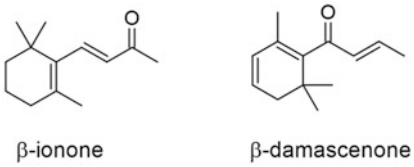


Fig. 8 Chemical structures of the major carotenoids in grapes. (Drawn by the authors with ChemSketch (ACD/Labs))

Fig. 9 Chemical structures of the representative C_{13}-norisoprenoids in grapes. (Drawn by the authors with ChemSketch (ACD/Labs))



interconversion between zeaxanthin and violaxanthin and between lutein and lutein 5,6-epoxide, named the xanthophyll cycle, is known to be involved in plant photoprotection processes. Thus, this conversion helps dissipate the excess of light energy in plant photosystems, shielding the photosynthetic apparatus. Carotenoid levels generally are higher during the first phase of berry growth, decreasing during berry ripening. Interestingly, only two carotenoids (zeaxanthin and antheraxanthin) have been reported to increase during berry ripening, reaching their maximum accumulation at harvest time.

Further modifications can occur to the end-products of the carotenoid pathway (neoxanthin and violaxanthin), which end up synthesizing compounds like abscisic acid, strigolactone, or flavor and aroma-related compounds like C_{13}-norisoprenoids (Lashbrooke et al. 2013). Within them, there are β-ionone and β-damascenone (Fig. 9), suggested to contribute to the grape floral and fruity aroma. In this regard, β-ionone is the direct cleavage metabolite of β-carotene, while 3-hydroxy-β-ionone

(a β-ionone derivative) represents the one of zeaxanthin and lutein. On the contrary, β-damascenone results from a multistep reaction starting with the dioxygenase cleavage of neoxanthin, producing the precursor ketodiol, which undergoes several acid-catalyze conversions (Sefton et al. 2011).

5 Vitamins and Other Compounds

Vitamins are of utmost importance because of their impact on human health and, above all, because we cannot synthesize them. Consequently, the only source for the human being is via dietary intake. Vitamins showing strong antioxidant potential are both water-soluble vitamins (such as vitamin C) and lipid-soluble (such as vitamin E). Ascorbic acid, or vitamin C (Fig. 10), is a major soluble antioxidant agent, supplying renewing protection against the damages inflicted by reactive oxygen species (ROS). In grapes, the content of vitamin C is relatively low compared to other fruits, like citrus. In this regard, it is known that grapevines use ascorbic acid as a precursor of tartaric acid, which belongs to the primary organic acids synthesized in grapes (together with malic acid), and the end-product of the irreversible ascorbic acid catabolism (Burbidge et al. 2021). The biosynthesis of ascorbate in mitochondria can follow different pathways. The most studied one is the Smirnoff-Wheeler pathway (known as the D-mannose/L-galactose pathway), which in grapes is highly active in immature green berries. It converts D-glucose-6-phosphate into GDP-D-mannose, then into GDP-L-galactose and L-galactono-1,4-lactone, to ultimately produce ascorbic acid. The alternative biosynthetic pathway originates from methyl-galacturonate, a pectin degradation product, that is converted into L-galactonate, re-joining the Smirnoff-Wheeler pathway intermediate L-galactono-1,4-lactone. This alternative route is suggested to be the most common one in grape berries after veraison (Melino et al. 2009).

Tocochromanols (comprised tocotrienols and tocopherols, collectively termed vitamin E) are a group of antioxidants known for preventing lipid oxidation (Fig. E). The structure of vitamin E holds a polar chromanol head group derived from the shikimate pathway and a hydrocarbon tail of isoprenoid origin. In this regard, tocotrienols and tocopherols are differentiated by containing a phytyl or a geranylgeranyl chain, respectively. Within each group of tocochromanols, molecules can be additionally distinguished in α-, β-, γ-, and δ- molecules based on the number and position of additional methyl groups present in the chromanol ring. Tocochromanols are primarily stored in the seed of grape berries (Fig. 11).

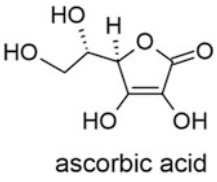


Fig. 10 Chemical structures of the ascorbic acid. (Drawn by the authors with ChemSketch (ACD/Labs))

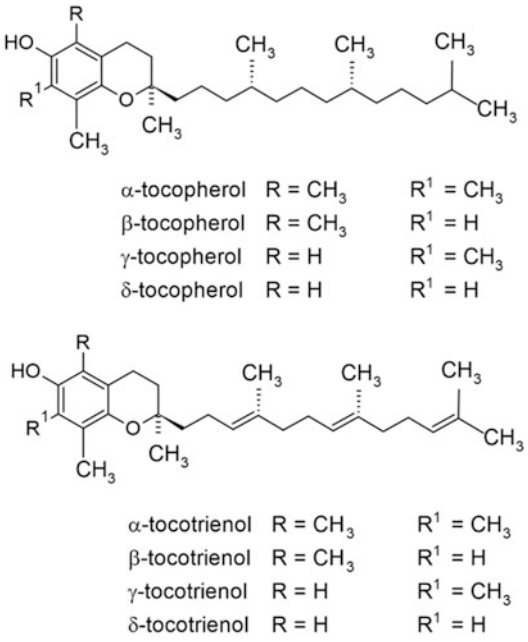


Fig. 11 Chemical structures of the tocochromanols. (Drawn by the authors with ChemSketch (ACD/Labs))

6 Genetic Resources and Extent of Genetic Diversity for Health-Related Compounds in Grapes

6.1 Diversity in the Muscadinia Subgenre

Grapes are part of the genus *Vitis*, a member of the Vitaceae plant family. According to morphological, anatomical, and cytological characteristics, species of the genus *Vitis* can be separated into two subgenera: *Muscadinia* ($2n = 2x = 40$) and *Vitis* (formerly known as *Euvitis*; $2n = 2x = 38$) (Aradhya et al. 2013). The natural occurrence of *Muscadinia* is narrowed to the Southeastern USA and Eastern Mexico (Aradhya et al. 2013), and it includes 2–3 species, of which only about 100 *V. rotundifolia* cultivars and interspecific hybrids have commercial interest (Yuzuak and Xie 2022). Fruits from these cultivars are commonly known as muscadine grapes, and they are grown for fresh fruit production and wine, juice, jam, and jelly elaboration. Muscadine grapes are suggested to have multiple health benefits, which result from their high content of phenolic compounds, namely anthocyanins and proanthocyanidins (Yuzuak and Xie 2022). Screening results suggest a high diversity in their levels of phenolic compounds, which are majorly found in grape skins and seeds (Pastrana-Bonilla et al. 2003). Through the analysis of 10 muscadine cultivars ('Carlos', 'Cowart', 'Early Fry', 'Fry', 'Ison', 'Late Fry', 'Noble', 'Paulk', 'Summit', and 'Supreme), Pastrana-Bonilla et al. (2003) highlighted the high concentration of ellagic acid, myricetin, quercetin,

kämpferol, and *trans*-resveratrol in muscadine grape skins, with concentration values ranging from 6.2 to 22.2, 1.8 to 19.6, 0.5 to 3.8, 0.2 to 3.8, and 0.1 to 0.2 mg/100 g, respectively. In seeds, epicatechin, catechin, and gallic acid dominate (Pastrana-Bonilla et al. 2003). Focusing on the ellagic acid and its precursors in eight muscadine cultivars ('Albermale', 'Carlos', 'Cowart', 'Doreen', 'Fry', 'Georgia Red', 'Nesbitt', and 'Noble'), Lee and Talcott (2004) reported a range of variation of 0.8 to 16.3, 0.7 to 11.5, and 58.7 to 190.0 mg/100 g, for free ellagic acid, ellagic acid glycosides, and total ellagic acid in the skin of ripe grapes, respectively. Regarding total anthocyanins, a wide range of variation in muscadine grapes has been indicated too, with an expected marked difference between black-berried cultivars (like 'Floriana' and 'Noble', with more than 10.0 mg/100 g dry weight) and white-berried cultivars (like 'Sweet Jenny', 'Watergate', and 'Welder', less than 0.03 mg/100 g dry weight) (Campbell et al. 2021). The overall high phenolic content of muscadine cultivars indicates their potential to develop new grape varieties with better fruit composition. However, hybrids between the *Muscadinia* and *Vitis* subgenera are generally infertile due to the discrepancy in the number of their chromosomes (Töpfer et al. 2011), limiting the usefulness of muscadine grapes in grapevine breeding programs.

6.2 Diversity Among Grape Vitis Species

The precise number of species belonging to the subgenus *Vitis* is challenging to estimate due to either hybridization between species or clinal variation within species. Thus, this subgenus is suggested to include between 60 and 70 species naturally occurring in Eastern Asia and North America and the only European grapevine species, *V. vinifera* (Péros et al. 2021). Comparative studies of the phenolic content and composition in ripe berries indicate significant differences among and within *Vitis* species. For example, the analysis of 147 grape accessions from 16 different *Vitis* species suggested that *V. rupestris* and *V. acerifolia* have the highest total phenolic concentration (20.2 mg/g and 19.2 mg/g fresh weight, respectively). On the other hand, *V. monticola*, *V. champinii*, and *V. labrusca* have the lowest concentration (3.8, 5.2, and 5.2 mg/g fresh weight) (Liang et al. 2012). Interestingly, these values are considerably higher than those reported for most of the *V. vinifera* L. cultivars analyzed by Liang et al. (2011) using the same methodology (1.4 mg/g fresh weight). A wide difference in anthocyanins content and composition between cultivars of different *Vitis* species has also been described. For example, delphinidin-derivatives might account for more than 60% of anthocyanins in species like *V. novaeangliae* and *V. champinii*, while it might be less than 20% in *V. monticola* (Liang et al. 2012). A wide range of variation for malvidin-derivatives (the dominant anthocyanins in *vinifera* cultivars (Sikuten et al. 2021) has been indicated too, accounting for up to 30% of total anthocyanins in species like *V. cinerea*, *V. palmata*, *V. vulpina*, *V. coignetii*, and *V. monticola*, to less than 2% (*V. novaeangliae*) (Liang et al. 2012). The concentration of total flavanols and flavonols also varies among species. Thus, *V. palmata* and *V. coignetii* have been indicated as some of the species with the highest content of flavanols, while

V. monticola and *V. champinii* have some of the lowest (Liang et al. 2012). Besides, *V. palmata*, *V. doaniana*, and *V. novaeangliae* are some of the species rich in flavonols, which contrasts with the low content reported in species like *V. monticola*, *V. champinii*, and *V. cinerea* (Liang et al. 2012). Substantial differences among *Vitis* species have also been indicated for stilbenes content and composition (Gabaston et al. 2020). As observed, non-*vinifera Vitis* species hold a wide range of variability for total phenolic content and individual phenolic compounds, which are of high interest for grape improvement. Nevertheless, some of these non-*vinifera* species carry certain off-flavors and undesirable aromas that, generally, are not positively perceived by consumers. As a result, the most common and realistic source of variability for developing novel grape cultivars with high fruit phenolic content is the one available within the *V. vinifera* species.

6.3 Diversity Among Wild and Cultivated *Vitis vinifera* Grapes

Most table and wine grape cultivars cultivated worldwide are part of the *V. vinifera* species, autochthonous of the Mediterranean Sea, southern and central Europe, northern Africa, and southwest and central Asia (Töpfer et al. 2011). Nowadays, two forms can still be found in these areas, the cultivated grapevine (*V. vinifera* subsp. *sativa* (or *vinifera*)) and its wild ancestor (*V. vinifera* subsp. *sylvestris*). Multiple sources indicate that the domestication process started in the Transcaucasian region during the early Neolithic Period (Myles et al. 2011). Genetic analyses based on different molecular markers indicate that the genetic diversity available in the *V. vinifera* subsp. *sativa* pool is highly structured, being linked to grape primary use (table or wine) and geographical origin (Bacilieri et al. 2013; Emanuelli et al. 2013; Laucou et al. 2018). Excluding novel grape cultivars achieved in breeding programs, these works indicate a major genetic division into four groups of cultivars: (i) table grapes from Eastern Mediterranean, Caucasus, Middle, and Far East countries, (ii) wine grapes from Western and Central Europe countries, (iii) wine grapes from the Balkans and Eastern Europe countries, and (iv) wine and table grapes from the Iberian Peninsula and the Maghreb. Further analyses indicate that this grouping correlates with significant differences in several grapevine reproductive and quality traits (Nicolas et al. 2016; Migicovsky et al. 2017). In this regard, differences in the phytochemical composition of *V. vinifera* subsp. *sativa* grapes between genetic groups have also been indicated (Teixeira et al. 2013; Sikuten et al. 2021).

In a comparative study between 11 *sylvestris* accessions from the eastern Adriatic region and three *sativa* cultivars ('Merlot', 'Plavac Mali', and 'Xinomavro'), Budic-Leto et al. (2018) indicated some significant differences in the phenolic content between subspecies, like those affecting the content of delphinidin-derivatives (higher in *sylvestris*), and of acylated anthocyanins (higher in *sativa*). Interestingly, the authors found two black-berried *sylvestris* accessions that lack acylated anthocyanins in berry skins, which agrees with previous findings by Revilla et al. (2012). The lack of acylated anthocyanins in berry skins is a rare characteristic in *sativa* cultivars, and it has only been observed in a few cultivars like the black-berried cultivar 'Pinot Noir' (and its offsprings and somatic variants) or the slightly colored

cultivar 'Gaglioppo' (Mattivi et al. 2006). Acylated anthocyanins in grapes not only increase phenolic fruit content but also stabilize the red color in wines during winemaking, leading to wines with a slight blue tint (Revilla et al. 2012). Regarding delphinidin-derivatives, delphinidin 3-*O*-glucoside was more abundantly found in *sylvestris* genotypes (on average, 22.5% of total anthocyanin content) than in the *sativa* cultivars 'Merlot', 'Plavac Mali', and 'Xinomavro' (on average, 10.3, 13.3, and 4.2%, respectively) (Budic-Leto et al. 2018).

Given its relevance for grape production, the phenolic content and composition of *V. vinifera* subsp. *sativa* cultivars have been screened in a vast number of genotypes. The analysis of the phenolic profiles of ripe berries from 344 cultivars by HPLC-MS revealed 36 phenolic compounds, including 16 anthocyanins, six flavonols, six flavanols, six hydroxycinnamic acids, and two hydroxybenzoic acids (Liang et al. 2011). Instead, the use of high-performance liquid chromatography coupled with triple quadrupole mass spectrometry analysis enabled the detection and quantification of 96 phenolic compounds, including the constitutive units of proanthocyanidins present in the skins of ripe grapes from 279 cultivars, providing highly detailed data on grape polyphenol composition (Pinasseau et al. 2017). These comprehensive analyses have suggested that polyphenol content varies among cultivars according to their primary use, with an overall higher level for all compounds (but hydroxycinnamic acids) in wine grapes than in table grapes (Liang et al. 2011). Nevertheless, the differences in phenolic content between cultivars of different use might be an indirect effect of their different behavior to factors affecting polyphenols biosyntheses, like precocity (Migicovsky et al. 2017) or the response to abiotic and biotic factors (Pinasseau et al. 2017). Considering only wine grapes, Sikuten et al. (2021) indicated that genetic and geographical background also affects the different phenolic content and composition between cultivars. These findings suggest that the evolution of the biosynthetic pathways leading to current grape phenolic content and composition might have been affected by regional environmental conditions, likely shaped by additional human-driven selection processes.

As expected, substantial differences in the total content of anthocyanins have been observed between grape cultivars of different berry colors. According to Liang et al. (2011), the average total anthocyanin content of black-berried cultivars is 30 times higher than the one found in pink-berried cultivars and 10 times higher than that of red-berried cultivars. As inferred from wine or table grapes studies, this effect is independent of the primary use of the cultivar. Regarding white-berried cultivars, only traces of anthocyanins have been detected from berry skin extracts in concentrations 5,000–60,000 times lower than in black-berried cultivars (Arapitsas et al. 2015). An interesting dietary source of anthocyanins is grapes from red-fleshed or 'teinturier' cultivars, which show an ectopic accumulation of anthocyanins in other plant organs, including berry flesh (Röckel et al. 2020). As a result, in proportion, grapes from 'teinturier' cultivars have a superior concentration of anthocyanins compared with white-fleshed cultivars (Kong et al. 2021; Röckel et al. 2020). Most of the available 'teinturier' cultivars derive from the French cultivar 'Teinturier', whose red-fleshed phenotype has been recently associated with the presence of a repetitive DNA element in the promoter of the *VviMybA1* gene (Röckel et al. 2020). In addition, other non 'Teinturier'-related red-fleshed cultivars

have been detected, like 'Gamay de Bouze', 'Gamay Fréaux' (somatic variants of 'Gamay'), and 'Pinot Teinturier' (a somatic variant of 'Pinot Noir') (Kong et al. 2021). Interestingly, cases of somatic variation causing red-fleshed berries have also been identified in table grape cultivars (Zhang et al. 2018), opening a promising sector for developing red-fleshed table grapes of higher nutritional quality.

On the contrary, berry skin color does not seem to significantly impact the content of other phenolic compounds (Liang et al. 2011). Flavanols are the major non-anthocyanin polyphenols of grapes, followed by flavonols. There is significant variability in the content of both groups of polyphenols, with up to a 100- and 176-fold variation between cultivars for flavanols and flavonols, respectively (Liang et al. 2011; Pinasseau et al. 2017). In this line, the cultivars 'Jampal' and 'Muscat St. Laurent' are suggested to be especially rich in flavanols, while 'Touriga Nacional' and 'Dornfelder' are in flavonols (Liang et al. 2011). The concentration of hydroxybenzoic and hydroxycinnamic acids has been indicated to vary widely between cultivars (Liang et al. 2011; Pinasseau et al. 2017), as does that of stilbenes (Gatto et al. 2008; Pinasseau et al. 2017). Regarding stilbene derivatives, the ratio between *trans*-resveratrol, *trans*-piceid, and *cis*-piceid also varies widely between cultivars (Gatto et al. 2008).

The characterization of different grape cultivars indicates that lutein and β-carotene are the major components of grape carotenoids, with significant differences between genotypes. As an example, the comparative analysis of nine grape cultivars grown under conventional conditions indicated a varying concentration ranging between 470.9 ± 46.0 and 825.0 ± 39.0 0 μg/kg for lutein in grape skins (for cultivars 'Aromat de Iaşi' and 'Feteascaregală', respectively), and between 229.6 ± 18.0 and 593.2 ± 35.0 μg/kg for β-carotene (for 'Napoca' and 'Muscat Ottonel' grape skins, respectively) (Bunea et al. 2012). Besides, different works indicate that vitamin content also varies significantly among grape cultivars. For example, vitamin C content varied between 11.2 ± 0.1 and 35.7 ± 0.3 mg/100 g in a group of nine Turkish table grape cultivars (Eyduran et al. 2015), a similar range to the one found in the analysis of 10 table grape cultivars from Syria (from 9.7 ± 1.2 to 30.9 ± 3.3 mg/100 g fresh weight) (Khalil et al. 2017). Regarding vitamin E content (α- and γ-tocopherols), the analysis of six table grape cultivars indicated a range of variation from 5.0 to 8.1 mg/kg fresh weight (values for 'Italia' and 'Muscat de Hambourg' cultivars, respectively) (Aubert and Chalot 2018). In the analysis of 15 wine and table grape cultivars, the total tocopherol content (α-, β-, γ-, and δ- tocopherols) was found to range from 29.2 ± 1.6 (in 'Muscat of Alexandria') to 102.7 ± 1.3 ('Syrah') mg/kg dry weight in bagasse (skin and pulp), and between 6.0 ± 1.1 ('Semillon') to 25.9 ± 1.7 ('Trakya Ilkeren') mg/kg dry weight in seeds (Tangolar et al. 2011). Following this work, the principal use of the cultivar (table or wine) does not significantly affect vitamin E content, which is dominated by α- and γ-tocopherols in grapes in both cases.

Another source of genetic variation affecting the content of health-related compounds in grapes is caused by somatic variation events. Spontaneous somatic mutations might cause changes in relevant phenotypic traits, deriving in clonal variation (Carbonell-Bejerano et al. 2019). To the interest of this chapter, different works report the identification of clones of some cultivars with significantly different phenolic content and composition (Ferrandino and Guidoni 2010; Muñoz et al. 2014; Pantelić et al. 2016; Royo et al. 2021). For example, the analysis of the

phenolic amount in ripe berries of four clones of 'Merlot' revealed a fold variation of 2.0, while that of four clones of 'Cabernet Franc' was set at 3.9 (Pantelić et al. 2016). Similarly, the maximum differences observed in the total phenolic content within 30 'Tempranillo Tinto' clones rose to 3.5-fold (Lemos et al. 2020). Following this work, the maximum fold variation in total anthocyanins was found to be 2.3, while one of the total flavonoids was 1.6.

7 Molecular Mapping Studies for Health-Related Compounds Content in Grapes

Understanding the genetic architecture of the accumulation of health-related compounds in grapevine berries implies identifying the number and location of the genomic regions (quantitative trait loci, QTLs) affecting trait variation, as well as their interaction. This information is essential for grapevine improvement, as it can be converted into practical knowledge for marker-assisted selection processes in modern breeding programs. Conventionally, QTL mapping is performed in populations segregating for the trait of interest through detecting significant associations between genetic markers segregation and trait phenotypic variation. This approach has been successfully applied over the past decades to reveal the genetic basis of traits related to the phenotypic determination of crop yield, grape quality, and adaptation mechanisms to abiotic and biotic factors. More recently, genome-wide association studies (GWAS) have also been implemented in different grapevine diversity panels to explore the genetic architecture of complex traits. Here, we focus on the works exploring the genetic basis of the biosynthesis of different health-related compounds in *V. vinifera* grape berries.

7.1 Anthocyanins Biosynthesis: The Berry Color Locus

Phenolic compounds are the most abundant and important health-related compounds in grape berries. The color of grape berries (and grape-derived products) is mainly associated with the content and composition of anthocyanins in berry skins. In addition, anthocyanin content clearly impacts the antioxidant potential of grapes. Given the relevance of berry color on fruit quality, the genetic basis of anthocyanins synthesis and accumulation in berry skins has been of great interest to the scientific community, and it can be reflected by the numerous works exploring this topic (see Ferreira et al. (2018) for a recent review). Early genetic studies in different *V. vinifera* progenies segregating for berry color indicated that this trait is mainly regulated by a single locus on chromosome 2, named *berry color locus*. The same locus was found when using quantitative data of total anthocyanins extracted from berry skins (Fournier-Level et al. 2009; Sun et al. 2020), which added evidence to support the role of anthocyanins on berry skin color determination. The oligonenic architecture of this trait has also been found in different GWAS (Myles et al. 2011; Migicovsky et al. 2017; Laucou et al. 2018), a procedure that screens functional variation in a

wider set, usually employing a panel of cultivars bearing most of the phenotypic variability existing for the target trait available at a species level.

Further molecular studies indicated that the *berry color locus* spans over a 200-kb region on chromosome 2, which encompasses a cluster of four MYB-type transcription factors: *VviMybA1*, *VviMybA2*, *VviMybA3,* and *VviMybA4* (Kobayashi et al. 2004; This et al. 2007; Walker et al. 2007; Fournier-Level et al. 2009). Different works have evidenced that both *VviMybA1* and *VviMybA2* are necessary for successfully accumulating anthocyanins in berry skins, indicating that white-berried cultivars (which lack anthocyanins in berry skins) only appear when these two genes are disrupted. This now widely accepted mechanism was first explored by Kobayashi et al. (2004) through the comparative study of the white-berried cultivar 'Italia' and its red-berried somatic variant 'Ruby Okuyama', which accumulates anthocyanins (mainly cyanidin-3-*O*-glucosides) in berry skins. Molecular and genetic studies revealed that the absence of anthocyanins in berry skins was tightly linked to the presence of a retrotransposon (called *Gret1*, from *Grape retrotransposon 1*) in the promoter of the *VviMybA1* gene sequence, which causes a non-functional allele. Thus, if this allele is found in homozygosis (as in 'Italia') it causes an absence of anthocyanins in berry skins, which derives in non-colored berries. This mechanism was then observed in a high number of cultivars of different origins and uses, which confirmed the role of *VviMybA1* on berry color variation at a species level (Lijavetzky et al. 2006; This et al. 2007). Later on, Walker et al. (2007) indicated that berry color loss is due to modifications in both *VviMybA1* and *VviMybA2*. Thus, besides the inclusion of the *Gret1* retrotransposon in *VviMybA1* first indicated by Kobayashi et al. (2004), two additional non-conservative mutations in *VviMybA2* are needed to inhibit the synthesis of anthocyanins in berry skins. Sequence analyses characterized these two mutations as (i) a single nucleotide polymorphism (SNP) causing a non-conservative amino acid substitution and (ii) a deletion of a dinucleotide that alters the reading frame in the functional *VviMybA2* allele. Considering that *VviMybA1* and *VviMybA2* are closely linked on the same chromosome region, they are usually considered part of a single MYB haplotype. This haplotype configuration determines a "white allele," which harbors non-functional alleles for both *VviMybA1* and *VviMybA2*, and two different "colored alleles," which harbor (i) *VviMybA1* functional and *VviMybA2* non-functional alleles, or (ii) *VviMybA1* and *VviMybA2* functional alleles (Ferreira et al. 2018). Functional analyses have indicated that both *VviMybA1* and *VviMybA2* are capable of inducing the *VviUFGT* gene expression (Walker et al. 2007). *VviUFGT* is the last gene of the phenylpropanoid pathway, and it has been suggested to be key for the biosynthesis of anthocyanins in berry skins and, consequently, berry pigmentation.

As previously indicated, genetic variations in *VviMybA1* have also been linked to the anthocyanin levels in the flesh of red-fleshed cultivars. After the analysis of a series of 'teinturier' cultivars, Röckel et al. (2020) associated the colored berry flesh phenotype with the presence of a 408-bp repetitive DNA element (called *Grapevine Color Enhancer*, GCE), found 338 bp upstream of the start codon in the *VviMybA1* gene promoter region. Following this work, GCE has been found with varying repetitions (two, three, and five) in different cultivars. Interestingly, the number of

GCE repeats correlates with the expression of *VviMybA1* and *VviUFGT* genes, as well as with berry flesh color intensity.

VviMybA3 sequence polymorphisms have also been significantly associated with berry color variation through a GWAS (Fournier-Level et al. 2009). However, it has been indicated that the cysteine-rich (CR) domain of the *VviMybA3* gene is truncated, deriving into a non-functional protein with poor functional evidence in the accumulation of anthocyanins in berry skins and berry skin color variation. However, it has been recently revealed that *VviMybA3* might contribute to regulating the accumulation of anthocyanins in berry flesh (Zhang et al. 2018). Finally, no associations between *VviMybA4* sequence polymorphisms and anthocyanins content variation have been described (Fournier-Level et al. 2009), sustaining the observed lack of functional activity of this transcription factor in grape berry skins (Walker et al. 2007). Beyond the *berry color locus*, Deluc et al. (2008) found that another MYB transcription factor (*VviMyb5b*, located in chromosome 6) can activate several genes of the general anthocyanins biosynthetic pathway, but not *VviUFGT*. Following this work, *VviMYB5b* has been suggested to participate in regulating proanthocyanidins biosynthesis in developing grape berries. In addition, Cardoso et al. (2012) tested the association between berry color and anthocyanins content variation and the genetic diversity detected in 15 genes that are not part of the *berry color locus* but putatively involved in the synthesis or transport of anthocyanins. Modeling results indicated some significant associations between the content of anthocyanins in berry skins and several polymorphisms found in three MYB transcription factors (*VviMYB11*, *VviMYBCC*, and *VviMYCB*). These results suggest that other genes not included in the *berry color locus* might influence anthocyanins biosynthesis mechanisms, which agrees with the additional minor QTLs for grape berry color and/or anthocyanins accumulation found in other studies (Sun et al. 2020).

7.2 Flavonols and Flavanols Biosynthesis: The Role of VviMYBF1, VviMybPA1, and VviMybPA2

The genetic basis of how flavonols are synthesized and accumulated in mature grape berries was explored by analyzing 170 individuals from a 'Syrah'×'Pinot Noir' population segregating for flavonol content and composition (Malacarne et al. 2015). According to this work, the *berry color locus* exerts a significant effect on flavonol content variation, indicating a common genetic control between flavonols and anthocyanins biosynthesis. This result agrees with previous findings that indicated that the metabolic reactions leading to the biosynthesis of these two groups of phenolic compounds are (at least) partially connected (Mattivi et al. 2006). Besides this QTL, other genomic regions associated with the fine-tuning of flavonol biosynthesis were identified, including a series of flavonol-specific QTLs that do not co-localize with genomic regions previously associated with anthocyanin biosynthesis or berry color variation. The authors found a region on chromosome 7 associated with kämpferol variation, which co-localizes with one MYB-type transcription factor gene, *VviMYBF1*. Interestingly, expression and functional analyses of *VviMYBF1* have

indicated that this transcription factor is a significant regulator of flavonols biosynthesis in developing grape berries (Czemmel et al. 2009).

Proanthocyanidins result from the polymerization of flavanol units. These compounds are secondary plant metabolites with multiple beneficial effects on human health, preferentially accumulated in berry skins and seeds. The genetic determinism of proanthocyanidin content in grape berries was first explored by considering two candidate genes: *VviMybPA1* and *VviMybPA2*. Using a series of transcriptome and functional studies, these two MYB-type transcription factors have been proposed as the central regulators of proanthocyanidins biosynthesis in grape berries (Bogs et al. 2007; Terrier et al. 2009; Carrier et al. 2013). In fact, different functional studies indicate that *VviMYBPA1* and *VviMybPA2* induce the transcription of key enzymes of the flavonoid, which triggers the biosynthesis of anthocyanins, flavonols, and proanthocyanidins in berries (Bogs et al. 2007; Terrier et al. 2009). Nevertheless, different genetic mapping studies have indicated that proanthocyanidin content might be under a highly more complex genetic control. For example, the analysis of 191 individuals from a segregating population obtained from a 'Syrah' × 'Grenache' cross revealed up to 43 and 103 QTLs for a series of skin and seed proanthocyanidins-content related traits (like total content or polymerization degree), respectively (Huang et al. 2012). These QTLs were virtually found on all chromosomes, suggesting a complex polygenic regulatory mechanism in berry skins and seeds. Interestingly, given the different numbers and positions of the QTLs identified for berry skins and seeds, results pointed out that the biosynthesis of proanthocyanidins in these two organs might be under different genetic regulation pathways. Based on a combination of QTL mapping and transcriptomic studies, Carrier et al. (2013) identified 20 genes in six chromosomes (1, 2, 3, 8, 14, and 17) as potential candidates to be involved in the proanthocyanidins biosynthetic pathway. Among them, three genes not previously associated with proanthocyanidins synthesis (*VviMybC2-L1*, *VviGAT-like*, and *VviCob-like*, in chromosomes 1, 3, and 17, respectively) were highlighted as promising candidates to be validated in subsequent works. Lastly, Huang et al. (2014) identified 21 eQTLs based on the transcript abundance of five downstream proanthocyanidins synthesis genes. Following this work, some of the more confident candidate eQTLs were related to genes linked to the general proanthocyanidins synthesis pathway, including *VviDFR* (in chromosome 18), *VviLAR1* (chromosome 1), or *VviLAR2* (chromosome 17).

7.3 Monoterpenes Biosynthesis: The 1-Deoxy-D-xylulose 5-Phosphate Synthase (DXS1) Activity

Other health-related compounds present in grape berries with antioxidant properties are monoterpenes. Grape monoterpenes have been classified into three categories: (i) free aroma compounds; (ii) free odorless polyols; and (iii) non-aromatic glycosidically bound forms of the monoterpenes (Mateo and Jiménez 2000). Given their contribution to the sweet and typical floral flavor of Muscat cultivars (Emanuelli et al. 2010), the genetic basis of the linalool, geraniol, and nerol content in grape

berries has been explored in multiple works. Different QTL mapping studies have indicated that muscat aroma (and/or geraniol, linalool, and nerol levels) is under an oligogenic control (Doligez et al. 2006; Battilana et al. 2009; Duchene et al. 2009; Wang et al. 2020). Recent GWAS have supported this genetic architecture using sets of grapevine cultivars of different origins (Migicovsky et al. 2017; Yang et al. 2017; Laucou et al. 2018; Guo et al. 2019; Liang et al. 2019). All these works indicate that geraniol, linalool, and nerol content in grapes is controlled by a key QTL on chromosome 5, which co-localizes with *VviDXS1* (Battilana et al. 2009; Duchene et al. 2009). This gene performs its action on the starting reaction of the plastidial MEP pathway (Battilana et al. 2009), whose functional role on monoterpene levels in grape berries was furtherly validated (Battilana et al. 2011). By association mapping, the causal SNP underlying this QTL has been identified (Emanuelli et al. 2010). Following this work, this SNP generates a non-neutral substitution present in the majority of Muscat-flavored cultivars. Beyond this locus, other minor QTLs have also been indicated by QTL mapping in different progenies (Doligez et al. 2006; Battilana et al. 2009; Duchene et al. 2009), suggesting the potential involvement of additional genomic regions to the final content of monoterpenes in grape berries.

7.4 Carotenoids Biosynthesis: What Gene Is Prominent?

As previously indicated, different carotenoids have been found in grapes, enhancing their nutritional value. Nevertheless, the information available on the genetic basis of their content in grape berries is limited. Using a comparative genomics approach, Young et al. (2012) identified up to 42 genes putatively involved in the grapevine carotenoid biosynthetic pathway, scattered on 16 chromosomes. More recently, after a genome-wide screening study, Leng et al. (2017) identified a total of 54 putative carotenoid metabolism-related genes, which were found to be distributed on 17 chromosomes. Altogether, these two works suggested a complex genetic architecture for grapevine carotenoid biosynthesis and provided a list of candidate genes likely involved in carotenoid accumulation. Nonetheless, specific genetic mapping analyses (either QTL mapping or GWAS, or both) are still needed to uncover the genetic basis underlying the accumulation of carotenoids in grape berries.

8 Grape miRNAs and Their Likely Impact on Human Health

Plant bioactive metabolites such as polyphenols, monoterpenes, carotenoids, and tocopherols are suggested to be responsible for the beneficial effects of consuming fruits and vegetables on human health. Plant microRNAs (miRNAs) are new bioactive molecules that migrate from plants to mammalian cells through dietary intake, regulating specific genes and pathways, and resulting in interesting candidates for dietary therapy approaches (Sanchita Trivedi et al. 2018; Li et al. 2021b).

miRNAs are endogenous, highly conserved, non-coding single-stranded RNAs. Specific miRNA genes code for these molecules, which consist of 18–24 nucleotides. At the post-transcriptional level, they can regulate gene expression by stimulating target messenger RNA (mRNA) degradation, translational repression, and chromatin modification of the corresponding target gene/s. Today, approximately half of the protein-coding mRNAs are thought to be influenced by miRNAs. miRNA gene silencing machinery is highly conserved in eukaryotes, even if notable differences exist between plants and animals concerning miRNA biogenesis, maturation, and mode of function (Achkar et al. 2016). These differences might contribute to the plant miRNA's high stability compared to their animal counterparts, explaining plant miRNA cross-kingdom abilities. Previous findings have revealed that plant miRNAs can survive adverse conditions such as low pH levels, high temperatures, ribonuclease and RNase activities, and digestive processes like food homogenization and absorption processes. In plants, adding a methyl group to the sugar present at the 3′-terminal nucleotide safeguards miRNAs from exonuclease degradation and 3′-uridylation. Moreover, the high plant miRNA GC content determines the absence of RNase digestion motifs (Yang et al. 2018). The nexus between Argonaute proteins or cofactors, such as high-density lipoproteins, with miRNA avoid miRNA decay, and combining miRNA with plant secondary metabolites creates a favorable setting that is able to inhibit RNAse activities. The packaging of plant miRNAs into microvesicles or exosomes protects their transportation before being absorbed via intestinal epithelial cells (van der Pol et al. 2012).

In recent years, plant miRNAs have attracted the attention of researchers for their cross-kingdom abilities and capability of regulating gene expression and cellular processes in humans through dietary intake. The first cross-kingdom study describing the detection of an exogenous plant miRNA (*miR168a*) in human serum was reported in 2012 (Zhang et al. 2012). Numerous other studies followed this pioneering work, and such studies can be found for diverse plant species like maize (Li et al. 2019), strawberry (Cavalieri et al. 2016), lettuce (Zhang et al. 2019), cabbage (Liang et al. 2014), spinach (Hou et al. 2018), and soybean (Chin et al. 2016). However, to our knowledge, only one cross-kingdom study has been conducted on grapes (Svezia et al. 2020). Following this work, authors proved the effects of grape miRNAs on human health by evaluating the cardioprotective role of cv. 'Sangiovese grape juice intake, using three different models: a murine model of myocardial infarction, murine coronary endothelial cells, and healthy human subjects. More specifically, they investigated the grape miRNAs impact on the natriuretic peptide system gene expression in murine coronary endothelial cells culture, representing the central auto-/paracrine signals of cardiac remodeling in infarcted patients. Remarkably, the survival of such cell lines was significantly linked to an intensified uptake of grapevine miRNAs. In the same work, for the first time, the authors detected the grape miRNAs in the human plasma of four healthy humans after seven days of 'Sangiovese' grape juice regular intake. Interestingly, the levels of two conserved plant miRNAs (*Vvi-miR159-3p* and *Vvi-miR166-3p*) increased in the bloodstream after 'Sangiovese' grape juice consumption.

To date, the ability of plant miRNAs to regulate key pathways in mammalian cells has been reported in an increasing number of studies. However, despite these promising results, works report contradicting evidence of cross-kingdom gene regulation by plant miRNAs, and the role of plant miRNAs on human metabolism regulation is still highly debated.

In plants, miRNAs are implicated in many functions related to growth and development, fruit ripening processes, signal transduction, response to abiotic and biotic stresses, and secondary metabolites modulation. In grapes, the publication of the genome sequences eased the discovery of miRNAs and their putative role in grapevine molecular regulatory networks. Many studies have evaluated the role of grape miRNA in developmental processes and stress responses. Belli Kullan et al. (2015) described the abundance of miRNAs in several grapevine tissues (including berries) at different developmental stages of the grapevine cultivar 'Corvina'. Accordingly, the distribution and abundance of miRNAs across samples are suggested to reflect the functional specificity of different organs, which ultimately aids in defining organ identity. Besides, Paim Pinto et al. (2016) compared the distribution of miRNAs in two grapevine cultivars ('Cabernet Sauvignon' and 'Sangiovese') collected in different vineyards and developmental stages. This work showed that both the developmental stage and cultivar influence miRNAs more than the vineyard. Chitarra et al. (2018) developed the first online database of grapevine miRNA candidates, called miRVIT. As a proof of concept study, this tool was successfully used to explore the response of 'Barbera' vines infected by *Flavescence dorée*, one of the most dangerous phytoplasma diseases.

The ingestion of plants with an added concentration of functional miRNAs in the diet has been recently suggested as a potential solution to face specific clinical situations (Sanchita Trivedi et al. 2018; Li et al. 2021a). However, further investigations and validations are required prior to this recommendation, and specific studies on many related topics (including how miRNAs are transported or absorbed by the human system) are needed to avoid any potential risks.

9 Applicability of Breeding Techniques in Grapevine Improvement

Improving crop traits, including nutritional value or quality, is critical to meet the growing population demands, which puts pressure on land use due to urbanization. Indeed, as a consequence of population growth and negative climate change effects on agricultural systems, there is a need to ensure food availability while maintaining (or increasing) food quality. In such context, crop breeding approaches (both conventional and modern activities) can improve crop productivity and quality features. However, some limitations have hindered grapevine breeding activities. Despite its importance, grapevine genetic improvement did not begin until the nineteenth century as a defense strategy against the arrival of powdery mildew, phylloxera, and downy mildew from North America (Töpfer et al. 2011). Although grapevine can be improved through several breeding techniques, these are difficult and time-consuming due to long generation cycles and the time required for selecting and

testing reliable progenies (juvenile stages). In addition, the grapevine is a highly heterozygous crop that exhibits inbreeding depression (Dalla Costa et al. 2019; Campos et al. 2021). Despite these limitations, grapevine improvement efforts could result in high-value products with high content of phenolic, aromatic, and vitaminic compounds that might provide beneficial effects linked to their antioxidant, anti-inflammatory, and anticancer properties, among others.

9.1 Conventional Breeding

Conventional grapevine breeding has focused on using strategies to counteract pathogens' negative effects and abiotic stressors on grapevine production. Chemical control is the most efficient way to deal with pathogens, but this approach has negative environmental, ecological, and sociological impacts. Thus, a way to reduce the use of chemicals is the development of novel resistant grapevine cultivars by the introgression of resistance R-loci from North American and Asian *Vitis* species into elite *vinifera* cultivars, developing more tolerant/resistant grapevine genotypes with high fruit quality (Töpfer et al. 2011; Pertot et al. 2017). To allow the preservation in time of the tolerant/resistant introgressed genes, gene pyramiding is used, aimed at incorporating more than one effective resistance gene (Pedneault and Provost 2016). This technique permits the parallel control of various pathogens or traits linked to grapevine resistance or tolerance. Moreover, these new genotypes might develop from multiple "back-crosses" steps with the elite cultivar, aimed to maintain a high percentage of the *V. vinifera* genome to preserve elite grape and wine quality properties. In this regard, several countries such as Austria, Chile, France, Germany, Hungary, Italy, Spain, and the USA have developed different wine and/or table grape breeding programs to obtain new cultivars with better yield and quality properties.

The development of improved genotypes holding genes from non-*vinifera* individuals while maintaining a high proportion of *V. vinifera* in their pedigree was speeded up by the use of Marker-Assisted Selection (MAS) (Pedneault and Provost 2016). In this sense, informative genetic markers associated with relevant grapevine features like berry size, acidity, color, aroma, and fruit ripening time have also been discovered in the grapevine. Most of these markers are reported in the "*Vitis* International Variety Catalogue – VIVC" website (https://www.vivc.de/docs/dataonbreeding/20220218_Table%20of%20Loci%20for%20Traits%20in%20Grapevine.pdf). Within them, some markers linked to genetic *loci* involved in the synthesis of metabolites with impact on the nutraceutical content of grapes can be found. For example, the available markers for berry color and muscat aroma (based on *VviMybA1* and *VviDXS1* gene sequences, respectively) could be used to trace the presence/absence of alleles related to increased color and increased aroma (so, to trace the presence/absence of antioxidant compounds) in breeding programs. The availability of these markers represents an interesting starting point to work on improving the nutraceutical properties of grapes via MAS.

The conventional breeding process comprises three stages: hybridization, line-fixation, and field trials. Plant breeding is an extensive logistical procedure requiring hundreds of thousands of vines in the hybridization and line-fixation stages;

however, the amount of plants is drastically lowered to small-selected vines in advanced breeding lines by the end of the breeding process. Therefore, considering the presented limitations, conventional breeding is hardly exploitable. Specifically, because of a prolonged lifecycle, severe inbreeding depression, and complex genetic control of enological properties, other genetic improvement strategies must be considered for *Vitis* (Gray et al. 2014; Litz et al. 2020).

9.2 The Potential of New Plant Breeding Techniques (NPBTs)

Over the past 15 years, biotechnological application in breeding programs has developed New Plant Breeding Technologies (NPBTs) that can modify only specific target DNA sequences without changing other regions (linkage drag) and without the need for long backcrossing stages, a limitation of conventional breeding techniques (Enfissi et al. 2021; Giudice et al. 2021). NPBTs are a new generation of techniques able to improve plant disease resistance, abiotic stress resilience, and added nutritional values (Lusser et al. 2012; Giudice et al. 2021). Compared with traditional breeding techniques, NPBTs increase the precision and accuracy of making changes in the genomes, reducing the time and efforts needed to produce novel crop cultivars that meet new requirements and potentially reducing the loss of important traits such as the biosynthesis of nutraceutical compounds. Several NPBTs make small modifications to the plant DNA and do not introduce foreign genes. Even considering the great potential of these techniques, and even when the modification is impossible to be differentiated from the ones triggered by spontaneous mutations or conventional breeding (Bortesi and Fischer 2015), their applicability has encountered legislative constraints.

The first developed NPBT strategy is called cisgenesis, which was proposed by Schouten et al. (2006). In this approach, one or more genes of interest can be isolated from a *Vitis* species that could potentially be utilized in conventional breeding and therefore transferred, maintaining its constitutional sequence composed of promoter, gene orientation, and terminator into the cultivar to be improved. There are a few examples of cisgenesis in grapevine: the proof of concept was demonstrated in 2016 using a recombinase system (Dalla Costa et al. 2016). This method allowed the production of genetically modified organism (GMO) plants that were easily selectable using conventional GMO approaches (i.e., antibiotics). Once the GMO plants containing the cisgene(s) were selected, through the activation of a recombinase system, it was possible to excise the transgene used for the detection of the transformed plants from the genomic DNA (Dalla Costa et al. 2016, 2020). Hence, the final products are plants containing only the cisgene(s), with their own promoter and terminator, without any exogenous sequence. Unfortunately, the transgene excision was never achieved completely, producing chimeric plants where the number of cells containing the transgene was limited but still present in the final product. Up to now, no real cisgenic grapevine plants have been obtained, and the difficulties related to this approach limit the application of this method.

The most recent and famous NPBT developed is genome editing (GE). GE, and specifically Clustered Regulatory Interspaced Short Palindromic Repeats (CRISPR)-Cas systems, can insert modifications in specific target DNA sequences without changing other sequences, avoiding the introduction of foreign DNA. Editing the genome is accomplished by applying three components: a protein with nuclease activity (e.g., Cas9, Cas12, Cas13, etc.), a single guide RNA (sgRNA) required to drive the Cas protein to the target sites, and a Protospacer Adjacent Motif (PAM), which is a short sequence upstream of the complementary DNA strand acting as a tag for the target site (Doudna and Charpentier 2014). Once the target sequence is identified, the endonuclease inserts a double-strand DNA (dsDNA) break and consequently stimulates the DNA repair pathway (Panda and Ray 2022), which are the two basic steps for taking advantage of the Cas systems in NPBTs. The CRISPR-Cas system could be used to obtain knock-out mutants, insert a DNA fragment using a donor vector through the homologous recombination system, base edit a target sequence to induce a specific mutation in regulatory sequences, and modify the epigenome (Vats et al. 2019; Khalil 2020; Giudice et al. 2021; Molla et al. 2021). Since the first application of CRISPR-Cas9, new advances have been achieved to improve its efficiency, versatility, and specificity (Gleditzsch et al. 2019; Giudice et al. 2021; Wang et al. 2021; Li et al. 2021b). Moreover, multiplex CRISPR-Cas9 gene editing can also be simultaneously accomplished using different gRNAs to edit a single gene and enhance editing efficiency (Najera et al. 2019), as demonstrated in a recent work where the targeting of *TAS4* and *MYBA7* genes efficiently prevented the accumulation of anthocyanins in the grapevine rootstock Milardet Grasset 101–14 (Sunitha and Rock 2020). On the other hand, genome editing applied to a gene belonging to the bZIP family allowed the over-accumulation of anthocyanin in plant tissue (Tu et al. 2022). More in detail, knocking out the *VvibZIP36* gene generated mutant plants, able to accumulate not only anthocyanin but also other secondary metabolites like naringenin chalcone, naringenin, dihydroflavonols, and cyanidin-3-*O*-glucoside. Although the potential of this technique is immense in woody plants, and specifically in grapevine, this strategy has many limitations, including the transformation method. This led laboratories to develop new delivery methods for plant systems. To date, methods such as the *Agrobacterium*-mediated transformation, nanoparticle platforms, biolistic transformation, or protoplast transfection can fulfill the role of delivering the DNA sequences encoding for Cas and sgRNA(s) into the host plant genome (Duan et al. 2021; Miller et al. 2021). In grapevine, the most used method to deliver CRISPR-Cas9 components into host cells still rely on *A. tumefaciens*-mediated transformation (Sandhya et al. 2020). This method includes the drawback of integrating the T-DNA into the plant genome (Lee and Gelvin 2008), leading to the production of edited plants.

The CRISPR-Cas approach has already been used and validated to enhance food quality, also from the nutraceutical point of view. For instance, the enhancement of provitamin A content in edible parts of relevant crops is considered an important issue and has been previously addressed (Zheng et al. 2021). Besides, genome editing techniques have been successfully applied to enhance nutrient availability

in cereals. For example, in a recent work, Ibrahim et al. (2021) used the CRISPR-Cas system to target genes implicated in the biosynthesis of phytic acid, responsible for the development of grains and involved in the bioavailability of iron and zinc in wheat, rice, and maize. Regarding the nutraceutical improvement of fruit crops, the CRISPR-Cas technology has been successfully used in 'Cavendish' banana cultivar to enhance β-carotene content by targeting the lycopene ε-cyclase (Kaur et al. 2020). Despite several limitations, these examples highlight the potentiality of CRISPR-Cas as a suitable strategy to improve grapevine nutraceutical features. Before being used, both genome editing and cisgenic approaches need the identification of suitable genes. Since these technologies work on one to a few target genes, exploiting traits under polygenic control would be challenging to achieve with the NPBTs. From the legislative perspective, plants containing exogenous DNA sequences (including marker genes necessary to confer resistance to selective chemicals for their early detection) are considered genetically modified organisms. In the case of plant species propagated by seeds, such as most of the herbaceous species (e.g., tomato, wheat, etc.), the elimination of the transgene containing T-DNA cassette could be achieved by co-transformation with different vectors and segregation of marker genes from the gene of interest in the progeny. Sadly, this approach is not applicable in vegetatively propagated plants characterized by long juvenile periods and highly heterozygous genomes. Despite these limitations, present-day plant genome editing applications remain the most versatile tool to improve the sustainability and nutraceutical features of several crops, including grapevine, but further studies are necessary to solve the abovementioned constraints.

10 Conclusion and Future Perspectives

Current grapevines show an extraordinary diversity of multiple health-related compounds, which contribute to their distinct antioxidant activities. Although some non-*vinifera* cultivars might not be a realistic source for commercial breeding programs, the high inter-cultivar variability available within the *V. vinifera* species might be a very relevant source to improve the nutritional quality of current grapes in the near future. In this process, selecting the most beneficial plant material is paramount, including, when available, the selection of the clone with the most adequate characteristics. This wide variability has aided in identifying the genetic architecture and the candidate genes responsible for phenotypic variation, which is the basis for detecting useful markers to assist breeding activities via marker-assisted selection. As summarized in this chapter, the leading role of anthocyanins in berry color and the importance of monoterpenes in muscat flavor (two traits with high relevance in breeding) led to multiple studies that ended up uncovering their genetic determinism. Then, this knowledge was efficiently used to design some genetic markers currently used in grapevine breeding activities. Nevertheless, the genetic architecture underlying the biosynthesis and accumulation of other grape health-related compounds is barely known. For example, the genetic basis of stilbene and stilbene derivatives (like resveratrol) accumulation in grape skins is unknown, although this compound is

known to have multiple positive biological effects on human health. Similarly, little is known about the genomic regions affecting the accumulation of other phenolics with high potential in preventing cardiovascular and other chronic diseases. The function of grapevine miRNA in regulating plant networks has emerged in recent years thanks to several studies, although their possible effect on humans by dietary intake of fresh grapes or juice is still unknown. This limited information is an opportunity for more intensive research studies to identify the genomic regions, genes, and miRNA involved in the biosynthesis of relevant grape health-related compounds. These studies will be benefited from the new generation of molecular tools, technologies, and computational resources available nowadays. In this line, novel genome editing technologies might be key to validating the putative role of some candidate genes on the accumulation of health-related compounds in grape berries. The information obtained from these studies is expected to speed up grapevine breeding activities in modern breeding programs, fostering the obtaining of new grape cultivars of higher nutritional value and antioxidant properties.

References

Achkar NP, Cambiagno DA, Manavella PA (2016) miRNA biogenesis: a dynamic pathway. Trends Plant Sci 21(12):1034–1044

Alston JM, Sambucci O (2019) Grapes in the world economy. In: Cantu D, Walker MA (eds) The grape genome. Springer Nature, Cham

Ammollo CT, Semeraro F, Milella RA, Antonacci D, Semeraro N et al (2017) Grape intake reduces thrombin generation and enhances plasma fibrinolysis. Potential role of circulating procoagulant microparticles. J Nutr Biochem 50:66–73

Aradhya M, Wang Y, Walker MA, Prins BH, Koehmstedt AM et al (2013) Genetic diversity, structure, and patterns of differentiation in the genus *Vitis*. Plant Syst Evol 299:317–330

Arapitsas P, Oliveira J, Mattivi F (2015) Do white grapes really exist? Food Res Int 69:21–25

Aubert C, Chalot G (2018) Chemical composition, bioactive compounds, and volatiles of six table grape varieties (*Vitis vinifera* L.). Food Chem 240:524–533

Bacilieri R, Lacombe T, Le Cunff L, Di Vecchi-Staraz M, Laucou V et al (2013) Genetic structure in cultivated grapevine is linked to geography and human selection. BMC Plant Biol 13(25):25

Banjanin T, Uslu N, Vasic ZR, Özcan MM (2020) Effect of grape varieties on bioactive properties, phenolic composition, and mineral contents of different grapevine leaves. J Food Process Preserv 45:s

Battilana J, Costantini L, Emanuelli F, Sevini F, Segala C et al (2009) The *1-deoxy-d-xylulose 5-phosphate synthase* gene co-localizes with a major QTL affecting monoterpene content in grapevine. Theor Appl Genet 118(4):653–669

Battilana J, Emanuelli F, Gambino G, Gribaudo I, Gasperi F et al (2011) Functional effect of grapevine *1-deoxy-D-xylulose 5-phosphate synthase* substitution K284N on Muscat flavour formation. J Exp Bot 62:5497–5508

Belli Kullan J, Pinto DLP, Bertolini E, Fasoli M, Zenoni S et al (2015) miRVine: a microRNA expression atlas of grapevine based on small RNA sequencing. BMC genomics 16:393

Bogs J, Jaffè FW, Takos AM, Walker AR, Robinson SD (2007) The grapevine transcription factor *VvMYBPA1* regulates proanthocyanidin synthesis during fruit development. Plant Physiol 143 (3):1347–1361

Bönisch F, Frotscher J, Stanitzek S, Rühl E, Wüst M et al (2014a) A *UDP-glucose: monoterpenol glucosyltransferase* adds to the chemical diversity of the grapevine metabolome. Plant Physiol 165:561–581

Bönisch F, Frotscher J, Stanitzek S, Rühl E, Wüst M et al (2014b) Activity-based profiling of a physiologic aglycone library reveals sugar acceptor promiscuity of family 1 UDP-glucosyltransferases from grape. Plant Physiol 166:23–39

Bontpart T, Marlin T, Vialet S, Guiraud J-L, Pinasseau L et al (2016) Two shikimate dehydrogenases, *VvSDH3* and *VvSDH4*, are involved in gallic acid biosynthesis in grapevine. J Exp Bot 67:3537–3550

Bortesi L, Fischer R (2015) The CRISPR/Cas9 system for plant genome editing and beyond. Biotechnol Adv 33:41–52

Budic-Leto I, Mucalo A, Ljubenkov I, Zdunic G (2018) Anthocyanin profile of wild grape *Vitis vinifera* in the eastern Adriatic region. Sci Hort 238:32–37

Bunea CI, Pop N, Babes AC, Matea C, Dulf FV et al (2012) Carotenoids, total polyphenols and antioxidant activity of grapes (*Vitis vinifera*) cultivated in organic and conventional systems. Chem Cent J 6:66

Burbidge CA, Ford CM, Melino VJ, Wong DCJ, Jia Y et al (2021) Biosynthesis and cellular functions of tartaric acid in grapevines. Front Plant Sci 12:643024

Campbell J, Sarkhosh A, Habibi F, Gajjar P, Ismail A et al (2021) Evaluation of biochemical juice attributes and color-related traits in muscadine grape population. Foods 10:1101

Campos G, Chialva C, Miras S, Lijavetzky D (2021) New technologies and strategies for grapevine breeding through genetic transformation. Front Plant Sci 12:767522–767522

Caputi L, Carlin S, Ghiglieno I, Stefanini M, Valenti L et al (2011) Relationship of changes in rotundone content during grape ripening and winemaking to manipulation of the 'peppery' character of wine. J Agric Food Chem 59:5565–5571

Carbonell-Bejerano P, Royo C, Mauri N, Ibanez J, Martinez-Zapater JM (2019) Somatic variation and cultivar innovation in grapevine. In: Morata A, Loira I (eds) Advances in grape and wine biotechnology. https://doi.org/10.5772/intechopen.86443

Cardoso S, Lau W, Eiras Dias J, Fevereiro P, Maniatis N (2012) A candidate-gene association study for berry colour and anthocyanin content in *Vitis vinifera* L. PLoS One 7(9):e46021

Carrier G, Huang YF, Le Cunff L, Fournier-Level A, Vialet S et al (2013) Selection of candidate genes for grape proanthocyanidin pathway by an integrative approach. Plant Physiol Biochem 72:87–95

Catalgol B, Batirel S, Taga Y, Ozer NK (2012) Resveratrol: French paradox revisited. Front Pharmacol 3:141

Cavalieri D, Rizzetto L, Tocci N, Rivero D, Asquini E et al (2016) Plant microRNAs as novel immunomodulatory agents. Sci Rep 6:25761

Chin AR, Fong MY, Somlo G, Wu J, Swiderski P et al (2016) Cross-kingdom inhibition of breast cancer growth by plant *miR159*. Cell Res 26(2):217–228

Chitarra W, Pagliarani C, Abbà S, Boccacci P, Birello G et al (2018) miRVIT: a novel miRNA database and its application to uncover *Vitis* responses to *Flavescence dorée* infection. Front Plant Sci 9:1034

Czemmel S, Stracke R, Weisshaar B, Cordon N, Harris NN et al (2009) The grapevine R2R3-MYB transcription factor *VvMYBF1* regulates flavonol synthesis in developing grape berries. Plant Physiol 151(3):1513–1530

Dalla Costa L, Piazza S, Campa M, Flachowsky H, Hanke M-V, Malnoy M (2016) Efficient heat-shock removal of the selectable marker gene in genetically modified grapevine. Plant Cell Tiss Org Cult (PCTOC) 124:471–481

Dalla Costa L, Malnoy M, Lecourieux D, Deluc L, Ouaked-Lecourieux F et al (2019) The state-of-the-art of grapevine biotechnology and new breeding technologies (NBTS). Oeno One 53:189–212

Dalla Costa L, Piazza S, Pompili V, Salvagnin U, Cestaro A et al (2020) Strategies to produce T-DNA free CRISPRed fruit trees via *agrobacterium tumefaciens* stable gene transfer. Sci Rep 10:1–14

Deluc L, Bogs J, Walker AR, Ferrier T, Decendit A et al (2008) The transcription factor *VvMYB5b* contributes to the regulation of anthocyanin and proanthocyanidin biosynthesis in developing grape berries. Plant Physiol 147:2041–2053

Doligez A, Audiot E, Baumes R, This P (2006) QTLs for Muscat flavor and monoterpenic odorant content in grapevine (*Vitis vinifera* L.). Mol Breed 18(2):109–125

Doudna JA, Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. Science 346:1258096

Drew DP, Andersen TB, Sweetman C, Møller BL, Ford C, Simonsen HT (2015) Two key polymorphisms in a newly discovered allele of the *Vitis vinifera* TPS24 gene are responsible for the production of the rotundone precursor α-guaiene. J Exp bot 67:erv491

Duan L, Ouyang K, Xu X, Xu L, Wen C et al (2021) Nanoparticle delivery of CRISPR/Cas9 for genome editing. Front Genet 12:673286

Duchene E, Butterlin G, Claudel P, Dumas V, Jaegli N et al (2009) A grapevine (*Vitis vinifera* L.) *deoxy-d-xylulose synthase* gene colocates with a major quantitative trait loci for terpenol content. Theor Appl Genet 118(3):541–552

Emanuelli F, Battilana J, Costantini L, Le Cunff L, Boursiquot J-M et al (2010) A candidate gene association study on Muscat flavor in grapevine (*Vitis vinifera* L.). BMC Plant Biol 10:241

Emanuelli F, Lorenzi S, Grzeskowiak L, Catalano V, Stefanini M et al (2013) Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. BMC Plant Biol 13:39

Enfissi EM, Drapal M, Perez-Fons L, Nogueira M, Berry HM et al (2021) New plant breeding techniques and their regulatory implications: an opportunity to advance metabolomics approaches. J Plant Physiol 258:153378

Eyduran SP, Akin M, Ercisli S, Eyduran E, Maghradze D (2015) Sugars, organic acids, and phenolic compounds of ancient grape cultivars (*Vitis vinifera* L.) from Igdir province of eastern Turkey. Biol Res 48:2

Falchi R, Wong DCJ, Yan Y, Savoi S, Gambetta GA, Castellarin SD (2019) The genomics of grape berry ripening. In: Cantu D, Walker MA (eds) The grape genome. Springer International Publishing, Cham, pp 247–274

FAO, OIV (2016) FAO-OIV Focus 2016. Table and dried grapes. Non-alcoholic products of the vitivinicultural sector intended for human consumption. Published by theFood and Agriculture Organization of the United Nationsand theInternational Organisation of Vine and Wine

Ferhi S, Santaniello S, Zerizer S, Cruciani S, Fadda A et al (2019) Total phenols from grape leaves counteract cell proliferation and modulate apoptosis-related gene expression in MCF-7 and HepG2 human cancer cell lines. Molecules 24:612

Ferrandino A, Guidoni S (2010) Anthocyanins, flavonols and hydroxycinnamates: an attempt to use them to discriminate *Vitis vinifera* L. cv 'Barbera' clones. Eur Food Res Technol 230:417–427

Ferreira V, Pinto-Carnide O, Arroyo-Garcia R, Castro I (2018) Berry color variation in grapevine as a source of diversity. Plant Physiol Biochem 132:696–707

Fournier-Level A, Le Cunff L, Gomez C, Doligez A, Ageorges A et al (2009) Quantitative genetic bases of anthocyanin variation in grape (*Vitis vinifera* L. ssp sativa) berry: a quantitative trait locus to quantitative trait nucleotide integrated study. Genetics 183:1127–1139

Gabaston J, Fonayet JV, Franc C, Waffo-Teguo P, De Revel G et al (2020) Characterization of stilbene composition in grape berries from wild *Vitis* species in year-to-year harvest. J Agric Food Chem 68:13408–13417

Gatto P, Vrhovsek U, Muth J, Segala C, Romualdi C et al (2008) Ripening and genotype control stilbene accumulation in healthy grapes. J Agric Food Chem 56:11773–11785

Giudice G, Moffa L, Varotto S, Cardone MF, Bergamini C et al (2021) Novel and emerging biotechnological crop protection approaches. Plant Biotechnol J 19:1495–1510

Gleditzsch D, Pausch P, Müller-Esparza H, Özcan A, Guo X et al (2019) PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. RNA Biol 16:504–517

Gray DJ, Li ZT, Dhekney SA (2014) Precision breeding of grapevine (*Vitis vinifera* L.) for improved traits. Plant Sci 228:3–10

Guo DL, Zhao HL, Li Q, Zhang GH, Jiang JF et al (2019) Genome-wide association study of berry related traits in grape (*Vitis vinifera* L.) based on genotyping-by-sequencing markers. Hortic Res 6:11

Gupta M, Dey S, Marbaniang D, Pal P, Ray S et al (2020) Grape seed extract: having a potential health benefits. J Food Sci Technol 57(4):1205–1215

Hall D, De Luca V (2007) Mesocarp localization of a bi-functional resveratrol/hydroxycinnamic acid glucosyltransferase of Concord grape (*Vitis labrusca*). Plant J 49:579–591

Herman F, Westfall S, Brathwalte J, Passinetti GM (2018) Suppression of presymptomatic oxidative stress and inflammation in neurodegeneration by grape-derived polyphenols. Front Pharmacol 9:867

Höll J, Vannozzi A, Czemmel S, D'Onofrio C, Walker AR et al (2013) The R2R3-MYB transcription factors *MYB14* and *MYB15* regulate stilbene biosynthesis in *Vitis vinifera*. Plant Cell 25: 4135–4149

Hou D, He F, Ma L, Cao M, Zhou Z et al (2018) The potential atheroprotective role of plant *MIR156a* as a repressor of monocyte recruitment on inflamed human endothelial cells. J Nutr Biochem 57:197–205

Huang YF, Doligez A, Fournier-Level A, Le Cunff L, Bertrand Y et al (2012) Dissecting genetic architecture of grape proanthocyanidin composition through quantitative trait locus mapping. BMC Plant Biol 12:30

Huang YF, Vialet S, Guiraud JL, Torregrosa L, Bertrand Y et al (2014) A negative MYB regulator of proanthocyanidin accumulation, identified through expression quantitative locus mapping in the grape berry. New Phytol 201(3):795–809

Ibrahim S, Saleem B, Naeem M, Arain S, Khan M (2021) Next-generation technologies for iron and zinc biofortification and bioavailability in cereal grains. Crop Pasture Sci 73:77–92

Ilc T, Werck-Reichhart D, Navrot N (2016) Meta-analysis of the core aroma components of grape and wine aroma. Front Plant Sci 7:1472

Kaur N, Alok A, Kumar P, Kaur N, Awasthi P et al (2020) CRISPR/Cas9 directed editing of *lycopene epsilon-cyclase* modulates metabolic flux for β-carotene biosynthesis in banana fruit. Metabol Engg 59:76–86

Khalil AM (2020) The genome editing revolution: review. J Genet Engg Biotechnol 18:68

Khalil S, Tello J, Hamed F, Forneck A (2017) A multivariate approach for the ampelographic discrimination of grapevine (*Vitis vinifera*) cultivars: application to local Syrian genetic resources. Genet Resour Crop Evol 64(8):1841–1851

Ko SH, Kim HS (2020) Menopause-associated lipid metabolic disorders and foods beneficial for postmenopausal women. Nutrients 12:202

Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-induced mutations in grape skin color. Science 304:982

Kong J, Wu J, Guan L, Hilbert G, Delrot S et al (2021) Metabolite analysis reveals distinct spatio-temporal accumulation of anthocyanins in two teinturier variants of cv. 'Gamay' grapevines (*Vitis vinifera* L.). Planta 253:84

Lashbrooke JG, Young PR, Dockrall SJ, Vasanth K, Vivier MA (2013) Functional characterisation of three members of the *Vitis vinifera* L. carotenoid cleavage dioxygenase gene family. BMC Plant Biol 13:156

Laucou V, Launay A, Bacilieri R, Lacombe T, Adam-Blondon AF et al (2018) Extended diversity analysis of cultivated grapevine *Vitis vinifera* with 10K genome-wide SNPs. PLoS One 13(2): e0192540

Lee L-Y, Gelvin SB (2008) T-DNA binary vectors and systems. Plant Physiol 146:325–332

Lee JH, Talcott ST (2004) Fruit maturity and juice extraction influences ellagic acid derivatives and other antioxidant polyphenolics in Muscadine grapes. J Agric Food Chem 52:361–366

Lemos AM, Machado N, Egea-Cortines M, Barros AI (2020) Assessment of quality parameters and phytochemical content of thirty 'Tempranillo' grape clones for varietal improvement in two distinct subregions of Douro. Sci Hort 262(109096):109096

Leng X, Wang P, Wang C, Zhu X, Li X et al (2017) Genome-wide identification and characterization of genes involved in carotenoid metabolic in three stages of grapevine fruit development. Sci Rep 7:4216

Li X-Y, Wen Y-Q, Meng N, Qian X, Pan Q-H (2017) Monoterpenyl glycosyltransferases differentially contribute to production of monoterpenyl glycosides in two aromatic *Vitis vinifera* varieties. Front Plant Sci 8:1226

Li M, Chen T, Wang R, Luo JY, He JJ et al (2019) Plant *MIR156* regulates intestinal growth in mammals by targeting the Wnt/β-catenin pathway. Am J Physiol-Cell Physiol 317:434–448

Li D, Yang J, Yang Y, Liu J, Li H et al (2021a) A timely review of cross-kingdom regulation of plant-derived microRNAs. Front Genet 12:613197

Li Y, Li W, Li J (2021b) The CRISPR/Cas9 revolution continues: from base editing to prime editing in plant science. J Genet Genomics 48:661–670

Liang Z, Owens CL, Zhong G-Y, Cheng L (2011) Polyphenolic profiles detected in the ripe berries of *Vitis vinifera* germplasm. Food Chem 129:940–950

Liang Z, Yang Y, Cheng L, Zhong GY (2012) Polyphenolic composition and content in the ripe berries of wild *Vitis* species. Food Chem 132:730–738

Liang G, Zhu Y, Sun B, Shao Y, Jing A et al (2014) Assessing the survival of exogenous plant microRNA in mice. Food Sci Nutr 2:380–388

Liang ZC, Duan S, Sheng J, Zhu S, Ni X et al (2019) Whole-genome resequencing of 472 *Vitis* accessions for grapevine diversity and demographic history analyses. Nat Commun 10:1190

Lijavetzky D, Ruiz-Garcia L, Cabezas JA, De Andres MT, Bravo G et al (2006) Molecular genetics of berry colour variation in table grape. Mol Genet Genomics 276(5):427–435

Lin J, Massonnet M, Cantu D (2019) The genetic basis of grape and wine aroma. Hortic Res 6:1–24

Litz RE, Pliego-Alfaro F, Hormaza JI (2020) Biotechnology of fruit and nut crops 2nd edition. CABI Publishing, Wallingfor, Oxfordshire, UK

Luan F, Mosandl A, Münch A, Wüst M (2005) Metabolism of geraniol in grape berry mesocarp of *Vitis vinifera* L. cv. Scheurebe: demonstration of stereoselective reduction, E/Z-isomerization, oxidation and glycosylation. Phytochemistry 66:295–303

Lusser M, Parisi C, Plan D, Rodríguez-Cerezo E (2012) Deployment of new biotechnologies in plant breeding. Nat Biotechnol 30:231–239

Maia M, Cavaco AR, Laureano G, Cunha J, Eiras Dias E et al (2021) More than just wine: the nutritional benefits of grapevine leaves. Foods 10:2251

Malacarne G, Costantini L, Coller E, Battilana J, Velasco R et al (2015) Regulation of flavonol content and composition in (Syrah × pinot noir) mature grapes: integration of transcriptional profiling and metabolic quantitative trait locus analyses. J Exp Bot 66(15):4441–4453

Martin DM, Aubourg S, Schouwey MB, Daviet L, Schalk M et al (2010) Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. BMC Plant Biol 10:226

Mateo JJ, Jiménez M (2000) Monoterpenes in grape juice and wines. J Chromatogr A 881:557–567

Mattivi F, Guzzon R, Vrhovsek U, Stefanini M, Velasco R (2006) Metabolite profiling of grape: flavonols and anthocyanins. J Agric Food Chem 54:7692–7702

Mattivi F, Vrhovsek U, Masuero D, Trainotti D (2009) Differences in the amount and structure of extractable skin and seed tannins amongst red grape varieties. Aust J Grape Wine Res 15:27–35

Melino VJ, Soole KL, Ford CM (2009) Ascorbate metabolism and the developmental demand for tartaric and oxalic acids in ripening grape berries. BMC Plant Biol 9:145

Migicovsky Z, Sawler J, Gardner KM, Aradhya MK, Prins BH et al (2017) Patterns of genomic and phenomic diversity in wine and table grapes. Hortic Res 4:17035

Milella RA, Gasparro M, Alagna F, Cardone MF, Rotunno S et al (2020) Gene expression signature induced by grape intake in healthy subjects reveals widespread beneficial effects on peripheral blood mononuclear cells. J Funct Foods 64:103705

Miller K, Eggenberger AL, Lee K, Liu F, Kang M et al (2021) An improved biolistic delivery and analysis method for evaluation of DNA and CRISPR-Cas delivery efficacy in plant tissue. Sci Rep 11:1–11

Molla KA, Sretenovic S, Bansal KC, Qi Y (2021) Precise plant genome editing using base editors and prime editors. Nat Plants 7:1166–1187

Morales-Castilla I, García de Cortázar-Atauri I, Cook BI, Lacombe T, Parker A et al (2020) Diversity buffers winegrowing regions from climate change losses. Proc Natl Acad Sci USA, vol 117, p 2864

Muñoz C, Gomez-Talquenca S, Chialva C, Ibáñez J, Martínez-Zapater JM et al (2014) Relationships among gene expression and anthocyanin composition of Malbec grapevine clones. J Agric Food Chem 62:6716–6725

Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F et al (2011) Genetic structure and domestication history of the grape. Proc Natl Acad Sci U S A 108(9):3457–3458

Najera VA, Twyman RM, Christou P, Zhu C (2019) Applications of multiplex genome editing in higher plants. Curr Opin Biotechnol 59:93–102

Nash V, Ranadheera S, Georgouspoulou EN, Mellor DD, Panagiotakos DB et al (2018) The effects of grape and red wine polyphenols on gut microbiota – a systematic review. Food Res Int 113: 277–287

Ndlovu T, van Jaarsveld F, Caleb OJ (2019) French and Mediterranean-style diets: contradictions, misconceptions and scientific facts – a review. Food Res Int 116:840–858

Nicolas SD, Péros JP, Lacombe T, Launay A, Le Paslier M-C et al (2016) Genetic diversity, linkage disequilibrium and power of a large grapevine (*Vitis vinifera* L) diversity panel newly designed for association studies. BMC Plant Biol 16:74

OIV (2019) 2019 statistical report on world Vitiviniculture. Organisation Internationale de la Vigne et du Vin

Paim Pinto DL, Brancadoro L, Dal Santo S, De Lorenzis G, Pezzotti M et al (2016) The influence of genotype and environment on small RNA profiles in grapevine berry. Front Plant Sci 7:1459

Panda G, Ray A (2022) Decrypting the mechanistic basis of CRISPR/Cas9 protein. Progr Biophys Mol Biol 172:60–67

Pantelić M, DabićZagorac D, Natić M, GasĬić U, Jović S et al (2016) Impact of clonal variability on phenolics and radical scavenging activity of grapes and wines: a study on the recently developed merlot and cabernet Franc clones (Vitis vinifera L.). PLoS ONE 11(10):e0163823

Pastrana-Bonilla E, Akoh CC, Sellappan S, Krewer G (2003) Phenolic content and antioxidant capacity of muscadine grapes. J Agric Food Chem 51:5497–5503

Pedneault K, Provost C (2016) Fungus resistant grape varieties as a suitable alternative for organic wine production: benefits, limits, and challenges. Sci Hortic 208:57–77

Péros JP, Cousins P, Launay A, Cubry P, Walker A et al (2021) Genetic diversity and population structure in Vitis species illustrate phylogeographic patterns in eastern North America. Mol Ecol 30:2333–2348

Pertot I, Caffi T, Rossi V, Mugnai L, Hoffmann C et al (2017) A critical review of plant protection tools for reducing pesticide use on grapevine and new perspectives for the implementation of IPM in viticulture. Crop Protec 97:70–84

Pinasseau L, Vallverdú-Queralt A, Verbaere A, Roques M, Meudec E et al (2017) Cultivar diversity of grape skin polyphenol composition and changes in response to drought investigated by LC-MS based metabolomics. Front Plant Sci 8:1826

Pintać D, Cetojević-Simin D, Berežni S, Orčić D, Mimica-Dukić N et al (2019) Investigation of the chemical composition and biological activity of edible grapevine (Vitis vinifera L.) leaf varieties. Food Chem 286:686–695

Rasines-Perea Z, Teissedre PL (2017) Grape polyphenols' effects in human cardiovascular diseases and diabetes. Molecules 22(1):68

Renaud S, de Lorgeril M (1992) Wine, alcohol, platelets, and the French paradox for coronary heart disease. Lancet 339:1523–1526

Revilla E, Carrasco D, Carrasco V, Benito A, Arroyo-Garcia R (2012) On the absence of acylated anthocyanins in some wild grapevine accessions. Vitis 51(4):161–165

Röckel F, Moock C, Braun C, Schwander F, Cousins P et al (2020) Color intensity of the red-fleshed berry phenotype of Vitis vinifera Teinturier grapes varies due to a 408 bp duplication in the promoter of VvmybA1. Genes 11:891

Royo C, Ferradás Y, Martínez-Zapater JM, Motilva MJ (2021) Characterization of Tempranillo negro (VN21), a high phenolic content grapevine Tempranillo clone, through UHPLC-QqQ-MS/MS polyphenol profiling. Food Chem 360(130049):130049

Sanchita Trivedi R, Asif MH, Trivedi PK (2018) Dietary plant miRNAs as an augmented therapy: cross-kingdom gene regulation. RNA Biol 15(12):1433–1439

Sandhya D, Jogam P, Allini VR, Abbagani S, Alok A (2020) The present and potential future methods for delivering CRISPR/Cas9 components in plants. J Genet Engg Biotechnol 18:1–11

Schmidlin L, Poutaraud A, Claudel P, Mestre P, Prado E et al (2008) A stress-inducible resveratrol O-methyltransferase involved in the biosynthesis of pterostilbene in grapevine. Plant Physiol 148:1630–1639

Schouten HJ, Krens FA, Jacobsen E (2006) Cisgenic plants are similar to traditionally bred plants: international regulations for genetically modified organisms should be altered to exempt cisgenesis. EMBO Rep 7:750–753

Sefton MA, Skouroumounis GK, Elsey GM, Taylor DK (2011) Occurrence, sensory impact, formation, and fate of damascenone in grapes, wines, and other foods and beverages. J Agric Food Chem 59:9717–9746

Siebert TE, Wood C, Elsey GM, Pollnitz AP (2008) Determination of rotundone, the pepper aroma impact compound, in grapes and wine. J Agric Food Chem 56:3745–3748

Sikuten I, Stambuk P, Tomaz I, Marchal C, Kontic JK et al (2021) Discrimination of genetic and geographical groups of grape varieties (*Vitis vinifera* L.) based on their polyphenolic profiles. J Food Comp Anal 102:104062

Singh CK, Liu X, Ahmad N (2015) Resveratrol, in its natural combination in whole grape, for health promotion and disease management. Ann N Y Acad Sci 1348:150–160

Sun L, Li S, Jiang J, Tang X, Fan X et al (2020) New quantitative trait locus (QTLs) and candidate genes associated with the grape berry color trait identified based on a high-density genetic map. BMC Plant Biol 20:302

Sunitha S, Rock CD (2020) CRISPR/Cas9-mediated targeted mutagenesis of *TAS4* and *MYBA7* loci in grapevine rootstock 101-14. Transgenic Res 29:355–367

Svezia B, Cabiati M, Matteucci M, Passino C, Pè ME et al (2020) Tuscany Sangiovese grape juice imparts cardioprotection by regulating gene expression of cardioprotective C-type natriuretic peptide. Eur J Nutr 59(7):2953–2968

Tangolar SG, Özogul F, Tangolar S, Yagmur C (2011) Tocopherol content in fifteen grape varieties obtained using a rapid HPLC method. J Food Comp Anal 24:481–486

Teixeira A, Eiras-Dias J, Castellarin SD, Gerós H (2013) Berry phenolics of grapevine under challenging environments. Int J Mol Sci 14:18711–18739

Terrier N, Torregrosa L, Ageorges A, Vialet S, Verriès C et al (2009) Ectopic expression of *VvMybPA2* promotes proanthocyanidin biosynthesis in grapevine and suggests additional targets in the pathway. Plant Physiol Biochem 149(2):1028–1041

This P, Lacombe T, Cadle-Davidson M, Owens CL (2007) Wine grape (*Vitis vinifera* L.) color associates with allelic variation in the domestication gene VvmybA1. Theor Appl Genet 114: 723–730

Töpfer R, Hausmann L, Harst M, Maul E, Zyprian E, Eibach R (2011) New horizons for grapevine breeding. Fruit veg Cereal Sci Biotechnol 5:79–100

Tu M, Fang J, Zhao R, Liu X, Yin W et al (2022) CRISPR/Cas9-mediated mutagenesis of *VvbZIP36* promotes anthocyanin accumulation in grapevine (*Vitis vinifera*). Hortic Res 9:uhac022

Tutino V, De Nunzio V, Milella RA, Gasparro M, Cisternino AM et al (2021) Impact of fresh table grape intake on circulating microRNAs levels in healthy subjects: a significant modulation of gastrointestinal cancer-related pathways. Mol Nutr Food Res 65:2100428

Ulaszewska M, Garcia-Aloy M, Vázquez-Manjarrez N, Soria-Florido MT, Llorach R et al (2020) Food intake biomarkers for berries and grapes. Genes Nutr 15:17

Unusan N (2020) Proanthocyanidins in grape seeds: an updated review of their health benefits and potential uses in the food industry. J Funct Foods 67(103861):103861

van der Pol E, Boing AN, Harrison P, Sturk A, Nieuwland R (2012) Classification, functions, and clinical relevance of extracellular vesicles. Pharmacol Rev 64(3):676–705

Vannozzi A, Dry IB, Fasoli M, Zenoni S, Lucchin M (2012) Genome-wide analysis of the grapevine stilbene synthase multigenic family: genomic organization and expression profiles upon biotic and abiotic stresses. BMC Plant Biol 12:130

Vats S, Kumawat S, Kumar V, Patil GB, Joshi T et al (2019) Genome editing in plants: exploration of technological advancements and challenges. Cell 8:1386

Walker AR, Lee E, Bogs J, McDavid DAJ, Thomas MR et al (2007) White grapes arose through the mutation of two similar and adjacent regulatory genes. Plant J 49(5):772–785

Wang H, Yan A, Sun L, Zhang G, Wang X et al (2020) Novel stable QTLs identification for berry quality traits based on high-density genetic linkage map construction in table grape. BMC Plant Biol 20:411

Wang T, Zhang C, Zhang H, Zhu H (2021) CRISPR/Cas9-mediated gene editing revolutionizes the improvement of horticulture food crops. J Agric Food Chem 69:13260–13269

Wightman JD, Heuberger RA (2014) Effect of grape and other berries on cardiovascular health. J Sci Food Agric 95:1584–1597

Wolkovich EM, García de Cortázar-Atauri I, Morales-Castilla I, Nicholas KA, Lacombe T (2018) From pinot to Xinomavro in the world's future wine-growing regions. Nat Clim Chang 8:29–37

Yang J, Xiao Y-Y (2013) Grape phytochemicals and associated health benefits. Crit Rev Food Sci Nutr 53(11):1202–1225

Yang Y, Shi Z, Reheman A, Jin JW, Li C et al (2012) Plant food delphinidin-3-glucoside significantly inhibits platelet activation and thrombosis: novel protective roles against cardiovascular diseases. PLoS One 7(5):e37323

Yang X, Guo Y, Zhu J, Niu Z, Shi G et al (2017) Genetic diversity and association study of aromatics in grapevine. J Am Soc Hortic Sci 142(3):225–231

Yang J, Elbaz-Younes I, Primo C, Murungi D, Hirschi KD (2018) Intestinal permeability, digestive stability and oral bioavailability of dietary small RNAs. Sci Rep 8:10253

Young PR, Lashbrooke JG, Alexandersson E, Jacobson D, Moser C et al (2012) The genes and enzymes of the carotenoid metabolic pathway in *Vitis vinifera* L. BMC Genomics 13:243

Yuzuak S, Xie D-Y (2022) Anthocyanins from muscadine (*Vitis rotundifolia*) grape fruit. Curr Plant Biol 30(100243):100243

Zhang L, Hou D, Chen X, Li D, Zhu L et al (2012) Exogenous plant *MIR168a* specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. Cell Res 22(1): 107–126

Zhang K, Liu Z, Guan L, Zheng Y, Jiu S et al (2018) Changes of anthocyanin component biosynthesis in 'summer black' grape berries after the red flesh mutation occurred. J Agric Food Chem 66:9209–9218

Zhang S, Sang X, Hou D, Chen J, Gu H et al (2019) Plant-derived RNAi therapeutics: a strategic inhibitor of HBsAg. Biomaterials 210:83–93

Zheng X, Kuijer HN, Al-Babili S (2021) Carotenoid biofortification of crops in the CRISPR era. Trends Biotechnol 39:857–860

Mango Nutrigenomics for Nutritional Security

Nimisha Sharma, Anil Kumar Dubey, and Ramya Ravishankar

Contents

Abstract

Nutrition is all about the study of food and the intake of food to stay healthy. As the community is developing, the world is also facing substantial challenges, such as malnutrition or hyperalimentation. Due to the imbalance of nutrition, chronic disease rates are also drastically increasing in the world. Furthermore, it is leading to high rates of obesity and diabetes in cities and villages. High rise in diet-related disorders such as obesity, cardiovascular diseases, diabetes etc., have resulted in seriousness regarding "Genomics of Nutrition" research worldwide. Therefore, the present global growth of the epidemic needs to be addressed through the promises of nutrigenomics. How genes and diet jointly may affect a person's health and risk of developing the illness could be well studied by nutrigenomics. The goal of nutrigenomics is to study the interaction of nutrients with the genome, proteome, and metabolome,

N. Sharma (✉) · A. K. Dubey
Division of Fruits and Horticultural Technology, ICAR-Indian Agricultural Research Institute, New Delhi, India

R. Ravishankar
Sun Valley Family Care, Peoria, AZ, USA

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_46

and it describes the affinity between these specific nutrients and nutrient regimes for quality health. A result of consolidated analyses that comprised fruit and vegetable utilization was connected to preventing coronary artery disease, cancer, and fatality, with similar results seen when fruits were scrutinized independently from vegetables. Mango (*Mangifera indica* L.) is one of the excellent tropical fruits in the world. Most of the mango tree parts are a rich reserve of bioactive compounds, which reside in leaves, bark, and fruit (pulp, peel, and stone). Contemporary studies have proven the presence of significant bioactive components of remedies in fruit waste parts like mango peel and kernel. Mangiferin, flavonoids, catechin, phenolic acids, and gallic acid are a few of the biologically active components contained in this fruit. Hence, the study of nutrigenomics in mango is very important to mitigate malnutrition and coronary diseases.

Keywords

Mango · Diet-gene interaction · Nutrigenomics · Nutrition · Food supplement

Abbreviations

BaCs	Bioactive compounds
BCO1	*Beta-carotene oxygenase 1*
BMI	Body mass index
DM	Dry matter
EA	Ellagic acids
FAO	Food and Agriculture Organization
FW	Fresh weight
GT	Gallotannins
IU	International units
MAB	Marker-assisted breeding
MTHFD1	Methylenetetrahydrofolate dehydrogenase
MTHFR	Methylenetetrahydrofolate reductase
NCDs	Noncommunicable diseases
NHANES	National Health and Nutrition Examination Survey
OECD	Organization for Economic Cooperation and Development
PKU	Phenylketonuria
PP	Polyphenols
QTL	Quantitative trait locus
RAE	Retinol activity equivalents
RDA	Recommended dietary allowances
SNP	Single nucleotide polymorphism
UFGT	Udp glucose: flavonoid-3-O-glucosyltransferase
US	United States
USDA	United States Department of Agriculture

1 Introduction

Nutrigenomics is defined as the relationship between nutrients, diet, and gene expression. It is a fascinating, upcoming field that explains the role of nutrition on gene expression. It brings together the science of bioinformatics, nutrition, molecular biology, genomics, epidemiology, and molecular medicine. The present chapter highlights the nutrigenomics research in mango. It includes the common outlook of nutrigenomics, relevant diseases, the role of single nucleotide polymorphism (SNP) in gene alteration, diet supplementation, and public consciousness in general and specifically in the fruit crop mango. It is very clear that with the accelerated changes in food habits and lifestyles, individuals are becoming more susceptible to diet-related disorders. Therefore, there is an imperative need to accelerate more research in this area so that the relationship between diet and health could be better understood and everyone could be benefitted from the genomic revolution (Neeha and Kinth 2013).

The new science of nutrigenomics teaches us what specific foods tell your genes and how food affects a person's genes and how a person's genes affect the way the body responds to food. What you eat directly determines the genetic messages your body receives. These messages, in turn, control all the molecules that constitute metabolism: the molecules that tell your body to burn calories or store them. "If you can learn the language of genes and control the messages and instructions, they give your body and metabolism, you can radically alter how food interacts with your body, lose weight, and optimize your health" (Hyman 2006; Aswini and Varun 2010). The nutritional phenotype of individuals could be more precisely accessed *via* the science of omics (Collins et al. 2003; German et al. 2011). It is cardinal to understand human health both by the role of diet in the fluctuating declaration of a genome and the role of genetics in the uncertain responses to diet (Gopalan 1992; Ghoshal et al. 2003; Ghosh 2010; Ghosh and Gorakshakar 2010). It is quite obvious and understandable that individuals respond distinctively to the same dietary consumption. The most manifest goal of actively preventing disease and improving the health of all individuals, of all ages, becomes nutrition's greatest golden chance and its strenuous provocation will be in establishing these basic relationships and implementing them (Gobard and Hurlimann 2009; Godbole et al. 2009). The comprehensive retort of metabolism to the quislings of the lifestyle and food choices, environmental oscillations, the status of nutrients, hereditary background, and epigenetic changes, within an individual at a discrete point in time (e.g., metabolic recessive) is possibly a susceptible and actionable reflection of nutritional and metabolic status (Zivkovic and German 2009). In India, some the diseases like epilepsy, type 2 diabetes, and neural tube defect disease are found to be associated with low nutrient uptake (Menon et al. 2010; Mohan et al. 2007a, b; Naushad et al. 2010). Further, nutritional status was observed in the Indian population (Rao 2001; Raj et al. 2007).

2 Need of Nutrigenomics Study for Fruit Crops Like Mango

It is a treasure house of nutrition and supports the economic stamina of the territory (FAO 2020) in newly industrialized countries, where tropical fruits are mostly grown. The acceleration in worldwide production of tropical fruits elevated rapidly from 5% to as high as 23% in 2019, based on the approximate anticipation that it will be one of the flourishing agricultural sectors (OECD/FAO 2020). Tropical fruits contain multitudinous health-promoting bioactive compounds such as phenolic acids, carotenoids, flavonoids, anthocyanins, vitamins, minerals, fatty acids, and fiber. These fruits comprise tremendously bountiful antioxidants, and phytochemicals rank them a predominant nutritional source with good medicinal properties (Rymbai et al. 2013; Acham et al. 2018; Laldinchhana Lalrengpuii et al. 2020) and assist in accomplishing nutritional security for an ever-increasing world population. As a result, to circumvent nutritional forfeiture and superior merchantability, fruits need to be harvested at the perfect phases. Postharvest deprivation is around 20–40% (Bantayehu and Alemaye 2019; Rajapaksha et al. 2021) because of their highly decaying character. Accommodating an integrated "multi-omics" perspective supports escalating the genomic knowledge and its implementation in developing improved cultivars. A study on 15,000 adults in several US communities emphasized that both high and low percentages of carbohydrates as part of routine uptake were associated with high mortality, although a 50% reduction in carbohydrate uptake lowers the risk (Seidelmann et al. 2018). Higher salt and low-quality fat due to the intake of more animal protein and decreased consumption of fruits and vegetables lead to higher mortality risk (Mazidi et al. 2019). Therefore, the nutritionist suggests to take more veggies in a routine manner as it maintains sound health. Increased intake of certain fruits like mango corresponds with a medley of beneficial health outcomes. It not only lowered the risk of obesity but also decreased chronic illness (Slavin and Lloyd 2012; Dreher et al. 2018).

Cardiovascular diseases and all-cause mortality (Aune et al. 2017) may be auxiliary with low risk with the excessive intake of fruits (apples, pears, and citrus fruits) and vegetables (green leafy and cruciferous vegetables). Studies on the intake of mango and its correlation with nutrient quality and health outcomes emphasize restricted information. Former research using NHANES 2001–2008 illustrated that the intake of mango in kids and adults corresponded with increased nutrients in comparison to those who did not consume mangoes. One cup (165 g) of raw mango give 100 kcal, 3 g dietary fiber, 277 mg potassium, 70 μg folate, DFE, 60 mg vitamin C, 90 μg vitamin A, RAE, 1060 μg beta-carotene, and 12 mg choline (USDA Database 2015). Therefore, it is considered as the supreme source of a healthy diet, although it is still less consumed in the United States. Papanikolaou and Fulgoni (2022) studied nutrient intakes, diet quality, and health results using data from NHANES 2001–2018 in children and adult mango consumers ($n = 291$; adults $n = 449$) compared with mango nonconsumers (children $n = 28{,}257$; adults $n = 44{,}574$). Children who consumed mangoes had a significantly lower daily intake of added sugar, sodium, and total fat, and a higher intake of dietary fiber, magnesium, potassium, total choline, vitamin C, and vitamin D, compared with

nonconsumers. In adults, mango consumers had significantly higher daily intakes of dietary fiber, magnesium, potassium, folate, vitamin A, vitamin C, and vitamin E and significantly lower intakes of added sugar and cholesterol, compared with nonconsumers. Mango consumption was also associated with better diet quality versus mango nonconsumers ($p < 0.0001$). Mango consumption in youngsters was associated with lower BMI *z*-scores, compared with nonconsumption. In adults, BMI scores, waist circumference, and body weight were significantly lower only in male mango consumers compared to mango nonconsumers. The key findings comprise a healthy nutrient pattern (more consumption of vegetables and fruits, whole grains, and less animal protein foods). The present data are affiliated with previously published data documenting numerous benefits associated with the inclusion of fruit within healthy dietary patterns.

3 Mango and Dietary Benefits

Mango (*Mangifera indica* L.) is a member of the Anacardiaceae family (more than 70 genera and 1000 varieties) and is known as the "king of fruits." It is widely consumed due to its exotic flavor, succulence, and sweet taste. Mango leaves, bark, and fruit (pulp, peel, and stone) are rich sources of bioactive compounds (BaCs). It contain proteins [0.36–0.40 g 100 g^{-1} fresh weight (FW) of pulp; 1.76–2.05% (w/w) of peel; 66.1 g kg^{-1} of kernel flour; and 3.0% (w/w) of leaves], vitamin A [0.135–1.872 mg 100 g^{-1} FW of pulp; 15.27 International Units (IU) in kernels; 1490 IU in leaves], vitamin C [7.8–172.0 mg 100 g^{-1} FW of pulp; 188–349 μg g^{-1} FW of peel; 0.17 g kg^{-1} DW of kernel flour; 53 mg 100 g^{-1} dry matter (DM) in leaves], carotenoids (0.78–29.34 μg g^{-1} FW of pulp; 493–3945 μg g^{-1} FW of peel), mangiferin (1690.4 mg kg^{-1} DM in peel; 4.2 mg kg–1 DW of kernel extract), phenolic compounds, dietary fiber (DF), carbohydrates, minerals, and other antioxidants known to have medicinal, nutritional, and industrial benefits. Certain diseases related to oxidative stress need several bioactive compounds due to their antioxidant properties. In mango fruit, only the pulp is used, while all other parts are relinquished that could be better utilized due to its therapeutic properties. Thus, there is an exigency to conduct research on all the bioactive constituents present in mango. These compounds not only provide substantial medical and nutritional properties but also have industrial applications, as well as role in defending the plant. The present-day, ascendant worldwide population leads to the dual challenges of nutritional insecurity and dietary disorders, engendering to health problems such as obesity, cancer, and cardiovascular disease (Clugston and Smith 2002). Mango encloses a complex mixture of antioxidants including xanthones and polyphenols whereby can help protect us from many diseases (Berardini et al. 2005). Mangoes are ingested as fresh fruit or are processed into enriched products such as nectar, puree, squash, or juice. In all instances, only the pulp is used, while the stones and peel are discarded, which results in a considerable waste of organic material. Details of nutrient composition of mango are given in Table 1.

Table 1 Nutrient composition of ripe mango pulp (Masibo and He 2008)

Component	Value (100 g–1 FW)	Component	Value (100 g–1 FW)
Water (g)	78.90–82.80	Calcium (mg)	6.10–12.80
Carbohydrate (g)	16.20–17.18	Phosphorus (mg)	5.50–17.90
Fiber (g)	0.85–1.06 Ash (g)	Iron (mg)	0.20–0.63
Protein (g)	0.36–0.40	Vitamin A (mg)	0.135–1.872
Fat (g)	0.30–0.53	Thiamin (mg)	0.020–0.073
Riboflavin (mg)	0.025–0.068	Ascorbic acid (mg)	7.80–172.00
Niacin (mg)	0.025–0.707	Tocopherol (mg)	1.12
Tryptophan (mg)	3.00–6.00	Lysine (mg)	32.00–37.00
Methionine (mg)	4.00	Lycopene (mg)	0.35

Polyphenolics: polyphenolics (PP) are the most widely distributed secondary metabolites and serve as the dominant antioxidant compounds. Gallic acid and six hydrolyzable tannins constituted 98% of the total polyphenolics. Other polyphenolics reported in mango pulp include flavonoids, xanthones, phenolic acids, and gallotannins (Berardini et al. 2005); m-hydroxybenzoic acid, vanillic acids, and apigenin (Masibo and He 2008); and hydroxybenzoic acid, m-coumaric acid, coumaric acid, ferulic acid, myricetin, mangiferin, catechins, epicatechin, quercetin, ellagic acids (EA), benzoic acid, and protocatechuic acid (Kim et al. 2007; Jasna et al. 2009; Gorinstein et al. 2011) (Table 2).

4 Therapeutic Potentials of Bioactive Compounds from Mango Fruit Wastes

Mango contains a congregation of several bioactive compounds and has been used as a significant herb in the traditional and Ayurvedic medicinal system for centuries (Shah et al. 2010). Bioactive components (mangiferin, flavonoids, catechin, phenolic acids, gallic acid, and gallic acid derivatives) of therapeutic nature were identified in the kernel and peel of mango. The medicinal value of these compounds has been assessed *in vitro* and minimal pre-clinically (Asif et al. 2016). The seed of mango is an important source of therapeutic health benefits (Momeny et al. 2012). Mango contains 20–60% seed of the whole and the kernel is 45–75% of the whole seed fruit (Maisuthisakul and Gordon 2009). Peel is a waste product of the mango processing industry. It consists of 15–20% of mango weight (Masibo and He 2008). Various important compounds are distributed in various concentrations in different parts of mango fruit like seed, peel, and pulp (Ignat et al. 2011;Ghuniyal 2015; Parvez 2016; Torres-León et al. 2016). Several polyphenols like alkylresorcinol, flavonols, gallotannins, xanthones, and benzophenone derivatives have been reported in mango fruit waste; peel and seed kernel antimicrobial (Gadallah and Fattah 2011; Shabani and Sayadi 2014), anti-inflammatory (Robles-Sánchez et al. 2009), antidiabetic (Ediriweera et al. 2017), analgesic, immune modulator (Sahu et al. 2007), and antioxidative (Khandare 2016), Wauthoz et al. 2007).

Table 2 Medicinal properties of polyphenols and vitamins present in mango

Polyphenols	Vitamins/carotenoids	Lupeol
1. Gallic acid acts as a substrate for polyphenol oxidase (PPO) in the pulp 2. Ellagitannins inhibit cancer cell proliferation *in vitro* 3. Gallic acid, mangiferin, myricetin, and flavan-3-ols (e.g., catechin and epicatechin) can prevent membrane lipid peroxidation and protect cells from Parkinson's disease 4. These antioxidants prevent coronary atherosclerosis lowering the levels of low-density lipoprotein cholesterol and triglycerides	1. Ascorbic acid is known to be a potent antioxidant that can eliminate reactive oxygen species (ROS) and maintain the membrane-bound antioxidant 2. Tocopherol, in its reduced state, act as a cofactor for the activity of a number of key enzymes and act as a substrate for oxalate and tartrate biosynthesis 3. Play roles in stress resistance and the synthesis of collagen, hormones, and neurotransmitters 4. Lower risk of degenerative diseases such as cancer, heart disease, inflammation, arthritis, immune system decline, brain dysfunction, and cataracts 5. b-carotene was the dominant carotenoid in mango plays a vital role against degenerative diseases such as cancer, cataracts, and muscular diseases, as well as neurological, inflammatory, and immune disorders	1. It is a pentacyclic triterpene. It possesses pharmacological properties, acting as a strong antioxidant, antimutagenic, anti-inflammatory, and antiarthritic agent 2. Lupeol also prevented 7, 12 dimethylbenz(a)anthracene-induced strand breaks in DNA, thereby reducing the incidence of tumors, lowering the tumor body burden, and causing a significant delay in the latency period of tumor appearance

Mango seed consists of about 29% shells, 68% kernel, and 3% testa (Diarra 2014). The composition of mango seed kernel varies according to different varieties (Barreto et al. 2008). Based on dry weight, 11% fat, 6.0% protein, 77% carbohydrate, 2.0% ash, and 2.0% crude fiber are the average composition of mango seed kernel. Mango seed kernel is high in minerals such as sodium, potassium, phosphorus, calcium, and magnesium (Sandhu et al. 2007). The mango seed kernel encompasses 52–56% unsaturated fatty acids and 44–48% saturated fatty acids (primary stearic acid). The mango seed kernel also comprehends a substantial amount of essential amino acids (lysine, leucine, and valine). Bioactive components that are embodied in mango kernel incorporate phytosterols (stigmasterol, campesterol, and also consists of vitamin K), sitosterol (β-sitosterols), tocopherols, and polyphenols (Soong and Barlow 2006).

The mango peel contains a prominent proportion of total dietary fiber (45–78%), distributed into soluble (16–28%) and insoluble (29–50%) fractions (Ajila et al. 2007). Furthermore, the mango peel also contains cellulose, hemicelluloses, pectin, lipids, proteins, carotenoids, and polyphenols. Apart from this, mango peel also has an appreciable amount of reducing sugars, and due to reducing sugars, mango peel is

also harnessed for the fermentation process, bioenergy, and various value-added products (Barreto et al. 2008). Sandhu et al. (2007) chronicled that mango peel comprises an elevated quantity of pectin (10–15%), and the soaking process before the extraction of pectin increases its yield to about 21%. The bioactive compounds or the polyphenolic connotations in 100 g of mango seed kernel comprise 20.7 mg tannin, 6.0 mg gallic acid, 12.6 mg coumarin, 7.7 mg caffeic acid, 20.2 mg vanillin, 4.2 mg mangiferin, 10.4 mg ferulic acid, 11.2 mg cinnamic acid, and 7.1 mg unknown compounds (Masibo and He 2008). Polyphenols include mangiferin pentoside, quercetin, syringic acid, and ellagic acid (Ajila et al. 2007).

5 Gene and Genomics to Study Nutrigenomics

Genomic data of both eukaryotes and prokaryotes are essential to understand the complete science of food. These genome sequences pave the way to get thorough knowledge about the composition of nutrients and their availability. Further, processing strategies and safety standards could be determined in a systematic way. It allows unlocking of a novel method to the post-genomic era that allows nutritionists to screen the genetic background and observe the omics as a whole (transcriptome, proteome, and metabolome). Genome science has resulted in development of new dietary strategies, targeted to supply the optimum nutrition for every person. These tools are the pivot of the ascending domain of nutrigenomics (Fig. 1a, b). The tools of genomics research come in to use to design markers from the candidate genes. It not only improves the health of humans but other living entities as well. Further, it accelerates the breeding efficiency of crops for quality traits and imparts better resistance against diseases (EFSA 2008; Kogel et al. 2010; Polesani et al. 2010). It allows breeders to design a genotype *in silico* based on the desired phenotype by the knowledge procured from recognizing the alleles at all places in a population. As far as the domains of food science, proteomics offers opportunities for discovering functional foods with metabolic effects. It is exceptionally pertinent in the research of proteins of fauna and flora for designer crop breeding (Agrawal et al. 2010), identifying new biomarkers (Pavlou and Diamandis 2010), and discovering therapeutic targets (Katz-Jaffe et al. 2009). By applying measurements of single biomarkers using traditional biochemical methods (Bakker et al. 2010), metabolomics is more helpful in identifying the complexities of metabolic regulation. Due to polyploidy in plants, genomes tend to be stupendous. Hence, because of the scale of the projects, the progression of entire plant genomes lags somewhat beyond other life forms. However, sequencing in a variety of agriculturally important genomes is complete or nearing completion and is forming the basis of a vital knowledge resource for food research. These genomes appraise from the frame of reference of production with agricultural nutritionists who are still struggling with the fundamental strategies for moving beyond the discovery of genes associated with essential nutrients to maximize their agricultural suitability and nutrient bioavailability under the Darwinian selection pressures that guided the organism's development (Lagaert et al. 2009). A genome of organisms represents the culmination of its evolutionary

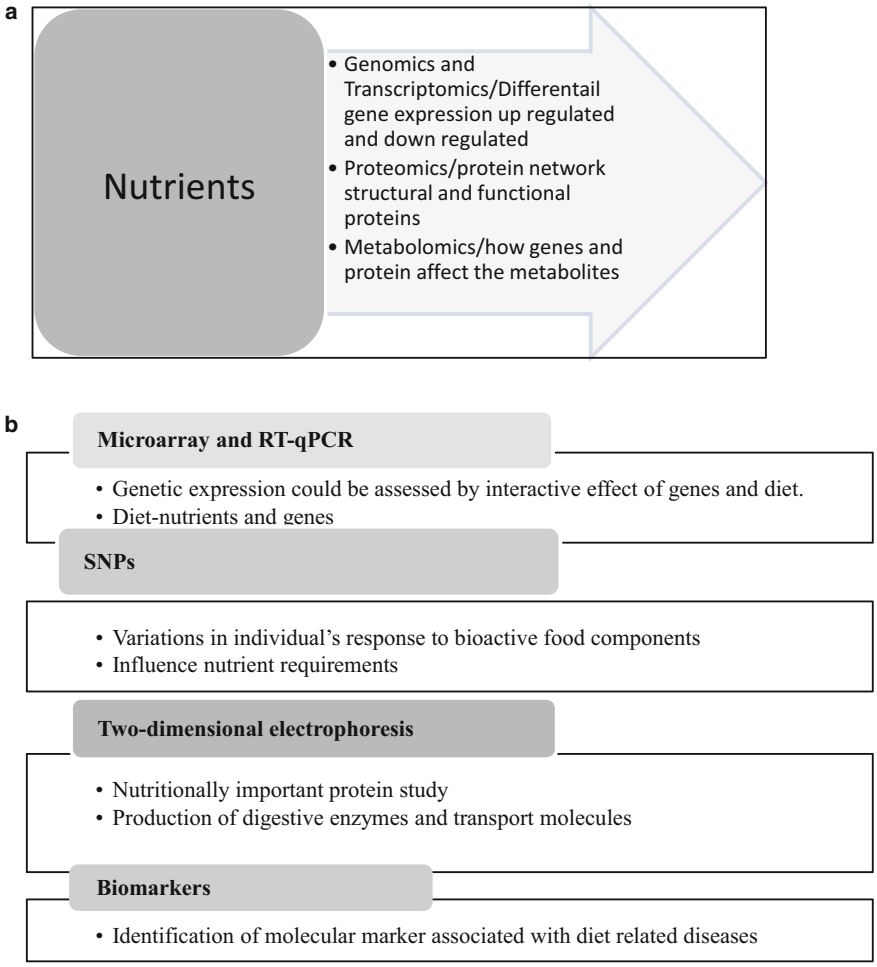


Fig. 1 (**a**, **b**) Conceptual nutrigenomics approach in mango for sound health. The model summarizes the proposed roles for various molecular tools to study the interaction among diet-nutrient and genes

history and the ensemble of genes emerging. Genetics could also influence nutrient and vitamin levels in individuals and changed gene expression. Hence, nutrigenetics emerges as a new science for unlocking the individual pedigree and their composition of nutrients. In mango, molecular markers like simple sequence repeats (SSRs) associated with fruit weight, width, volume, total soluble solid (TSS), titrable acidity, ascorbic acid, and total sugars, additionally reducing sugars, and succors will facilitate screening for varieties/seedlings with preferable fruit traits. Spongy tissue is a consequential physiological disorder influencing the palatable standard of mango drupes, producing a metabolic profile consisting of stress-related and flavor-suppressing metabolites that differ among different

stages of the fruit (Ajila et al. 2007; Masibo and He 2008). Genes also affect the absorption, transportation, and activation of nutrients and vitamins. There are studies on markers like single nucleotide polymorphisms (SNPs) which affect vitamin availability and cause deficiency (Rubab et al. 2022). Recently transcriptome-based grouping study showed the impact of mango fruit as dietary intake on cardiometabolic health that appears to have interindividual variability (Keathley et al. 2022). A study on the expression of genes involved in carotenoids and anthocyanins during ripening in fruit peel of green-, yellow-, and red-colored mango cultivars was carried out (Karanjalker et al. 2018). Various genes involved in bioactive and metabolites' production in mango like *UDP-glucose: flavonoid-Oglycosyl-transferase* (UFGT), *dihydroflavonol* (UFGT), *dihydroflavonol 4reductase*, and *anthocyanin synthase* are responsible for flavonoid synthesis (Karanjalker et al. 2018). Similarly, *lycopene-β-cyclase* is responsible for carotenoid synthesis. The gene *MiUFGT2* synthesizes cyanidine-3-O-monoglucosides and peonidin-3-O-glucosides bioactive compounds (Bajpai et al. 2018). Transcriptomes (a set of RNA) are transcribed at a cellular level. Functional products derived through RNA not only affect physiological functioning but also impart knowledge about disease progression (Passos 2015). Although, gene response may be variable, environmental interaction also change the expression of the cell. Some nutrition composition improves the health of humans *via* transcriptomic modulation. Further, "Omics" science could identify new therapeutic targets relevant to variable conditions (Chambers et al. 2019). The transcriptomic analyses provide insights to identify new metabolic pathways affected by mango consumption in individuals who responded to the intervention (Ducheix et al. 2018; Stefania et al. 2021). A network of pathways such as hydrogen peroxide, cofactor catabolic and metabolic processes, gases (oxygen, carbon dioxide) transport, and transcriptional regulation (by *RUNX1* and *TP53*) play a key role in this process. Metabolic pathways also modulated by mango consumption such as *RUNX1* have been demonstrated to be one of the most frequently mutated genes in several hematological malignancies (Sood et al. 2017). Further, *TP53* was found as expressed protein variant in human carcinoma (Khurana et al. 2016). Insulin resistance, diabetes, cardiovascular disease, and chronic diseases like cancer and kidney illness are regulated by hydrogen peroxide metabolism (Lismont et al. 2019). Previous studies with mango extracts have demonstrated beneficial health effects related to these conditions (Awodele et al. 2015; Fomenko and Chi 2016; Imran et al. 2017). Differential gene expression includes *TNFAIP3, API5*, and *TAL1*, etc., as up- and downregulated genes, that regulate the physiological processes. Based on these mechanistic findings, it appears that mango consumption could have a beneficial cardiometabolic effect. Moreover, a reduction in blood pressure following mango consumption is observed (Fang et al. 2018). Research on mangiferin component shows key inflammatory pathways involved in cancer progression (Gold-Smith et al. 2016; Imran et al. 2017; Piccolo et al. 2022). These mechanistic findings provide partial evidence supporting the anticancer potential of mangos.

6 Research Gaps and Future Prospects

The propagation of tropical fruit trees for ameliorating fruit traits is intricate due to several constraints like long gestation period, heterozygous nature, variable embryonic nature, and lack of high-quality genome sequences (Mathiazhagan et al. 2021). New advanced molecular tools supplement conventional breeding efforts. Numerous genomics strategies have recently progressed to accommodate and compliment to traditional breeding methods. DNA-based markers associated with fruit burgeoning and fruit quality characteristics were identified in perennial fruit crops. Furthermore, it could be utilized in association mapping.

7 Conclusion

With the availability of genome sequences of fruit crops, identification of SNP variants/Indels, QTLs, functional genes, etc., could be utilized in quality fruit production. Hence, the fruit superiority was accredited through multi-omics perspectives. Moreover, the recognition and measurements of transcripts involved in sugar-starch metabolism, fruit development and ripening, genomic selection (GS), and genetic modifications *via* transgenics have paved the way for studying gene function and developing varieties with improved quality traits of fruit crops by overcoming long breeding cycles.

References

Acham IO, Ahemen S, Ukeyima MT, Girgih AT (2018) Tropical fruits: bioactive properties and health promoting benefits in chronic disease prevention and management. Asian Food Sci J3:1–13

Agrawal GK, Jwa NS, Lebrun MH, Job D, Rakwal R (2010) Plant secretome: unlocking secrets of the secreted proteins. Proteomics 10(4):799–827

Ajila CM, Naidu KA, Bhat SG, Rao UP (2007) Bioactive compounds and antioxidant potential of mango peel extract. Food Chem 105(3):982–988

Asif A, Farooq U, Akram K, Hayat Z, Shafi A, Sarfraz F, Sidhu MA, Rehman HU, Aftab S (2016) Therapeutic potentials of bioactive compounds from mango fruit wastes. Trends Food Sci Technol 53:102–112

Aswini YB, Varun S (2010) Genetics in public health: rarely explored. Indian J Hum Genet 16:47–54

Aune D, Giovannucci E, Boffetta P, Fadnes LT, Keum N, Norat T, Greenwood DC, Riboli E, Vatten LJ, Tonstad S (2017) Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality—a systematic review and dose-response meta-analysis of prospective studies. Int J Epidemiol 46(3):1029–1056

Awodele O, Adeneye AA, Aiyeola SA, Benebo AS (2015) Modulatory effect of *Mangifera indica* against carbon tetrachloride induced kidney damage in rats. Inter Toxicol 8(4):175

Bajpai A, Khan K, Muthukumar M, Rajan S, Singh NK (2018) Molecular analysis of anthocyanin biosynthesis pathway genes and their differential expression in mango peel. Genome 61(3):157–166

Bakker GC, Van Erk MJ, Pellis L, Wopereis S, Rubingh CM, Cnubben NH, Kooistra T, Van Ommen B, Hendriks HF (2010) An antiinflammatory dietary mix modulates inflammation and

oxidative and metabolic stress in overweight men: a nutrigenomics approach. Am J Clin Nutr 91 (4):1044–1059
Bantayehu M, Alemaye M (2019) Estimation of pre and postharvest losses of tropical fruits in Ethiopia estimation of pre and postharvest losses of tropical fruits in Ethiopia 47. Int J Postharvest Technol Innov 6:46–56
Barreto JC, Trevisan MT, Hull WE, Erben G, De Brito ES, Pfundstein B, Würtele G, Spiegelhalder B, Owen RW (2008) Characterization and quantitation of polyphenolic compounds in bark, kernel, leaves, and peel of mango (*Mangifera indica* L.). J Agric Food Chem 56 (14):5599–5610
Berardini N, Knödler M, Schieber A, Carle R (2005) Utilization of mango peels as a source of pectin and polyphenolics. Innov Food Sci Emerg Technol 6(4):442–452
Chambers DC, Carew AM, Lukowski SW, Powell JE (2019) Transcriptomics and single-cell RNA-sequencing. Respirology 24(1):29–36
Clugston GA, Smith TE (2002) Global nutrition problems and novel foods. Asia Pacific J Clin Nutr 11:S100–S111
Collins YF, McSweeney PL, Wilkinson MG (2003) Lipolysis and free fatty acid catabolism in cheese: a review of current knowledge. Int Dairy J 13(11):841–866
Diarra SS (2014) Potential of mango (*Mangifera indica* L.) seed kernel as a feed ingredient for poultry: a review. World's Poul Sci J 70(2):279–288
Dreher A, Fuchs A, Parks B, Strange AM, Tierney MJ (2018) Apples and dragon fruits: the determinants of aid and other forms of state financing from China to Africa. Int Stud Quart 62 (1):182–194
Ducheix S, Peres C, Härdfeldt J, Frau C, Mocciaro G, Piccinin E, Lobaccaro JM, De Santis S, Chieppa M, Bertrand-Michel J, Plateroti M (2018) Deletion of stearoyl-CoA desaturase-1 from the intestinal epithelium promotes inflammation and tumorigenesis, reversed by dietary oleate. Gastroenterology 155(5):1524–1538
Ediriweera MK, Tennekoon KH, Samarakoon SR (2017) A review on ethnopharmacological applications, pharmacological activities, and bioactive compounds of *Mangifera indica* (Mango). Evidence-Based Complem Altern Med 6949835, pp. 1–24 https://doi.org/10.1155/2017/6949835
EFSA (Eur. Food Saf. Auth.) Panel Working Group on Animal Feeding Trials (2008) Safety and nutritional assessment of GM plants and derived food and feed: the role of animal feeding trials. Food Chem Toxicol 46:2–70
Fang C, Kim H, Barnes RC, Talcott ST, Mertens-Talcott SU (2018) Obesity-associated diseases biomarkers are differently modulated in lean and obese individuals and inversely correlated to plasma polyphenolic metabolites after 6 weeks of mango (*Mangifera indica* L.) consumption. Mol Nutr Food Res 62(14):1800129
FAO (2020) Major tropical fruits market review. FAO, Rome
Fomenko EV, Chi Y (2016) Mangiferin modulation of metabolism and metabolic syndrome. BioFact 42(5):492–503
Gadallah MG, Fattah AA (2011) The antibacterial effect of mango seed kernel powder in minced beef during refrigerated storage. World J Dairy Food Sci 6(2):219–228
German JB, Zivkovic AM, Dallas DC, Smilowitz JT (2011) Nutrigenomics and personalized diets: what will they mean for food? Annu Rev Food Sci Technol 2:97
Ghosh D (2010) Personalised food: how personal is it? Genes Nutr 5:51–53
Ghosh K, Gorakshakar A (2010) Integration of modern genetic knowledge and technology into public health in India. Indian J Hum Genet 16:45–46
Ghoshal S, Pasham S, Odom DB, Furr HC, McGrane MM (2003) Vitamin A depletion is associated with low phosphoenolpyruvate carboxykinase mRNA levels during late fetal development and at birth in mice. J Nutr 33:2131–2136
Ghuniyal J (2015) Ethanomedical, chemical, pharmacological, toxicological properties of *Mangifera indica*: a review. Int J Pharma Res Rev 4(10):51–64

Gobard B, Hurlimann T (2009) Nutrigenomics for global health: ethical challenges for underserved populations. Curr Pharmacogenom Personaliz Med 7(3):205–214
Godbole K, Deshmukh U, Yajnik C (2009) Nutri-genetic determinants of neural tube defects in India. Indian Pediat 46:467–475
Gold-Smith F, Fernandez A, Bishop K (2016) Mangiferin and cancer: mechanisms of action. Nutrients 8(7):396
Gopalan C (1992) The contribution of nutrition research to the control of under nutrition: the Indian experience. Annu Rev Nutr 12:1–17
Gorinstein S, Poovarodom S, Leontowicz H, Leontowicz M, Namiesnik J, Vearasilp S, Haruenkit R, Ruamsuke P, Katrich E, Tashma Z (2011) Antioxidant properties and bioactive constituents of some rare exotic Thai fruits and comparison with conventional fruits: *in vitro* and *in vivo* studies. Food Res Int 44(7):2222–2232
Hyman M (2006) Ultrametabolism: the simple plan for automatic weight loss. Simon and Schuster
Ignat I, Volf I, Popa VI (2011) A critical review of methods for characterization of polyphenolic compounds in fruits and vegetables. Food Chem 126(4):1821–1835
Imran M, Arshad MS, Butt MS, Kwon JH, Arshad MU, Sultan MT (2017) Mangiferin: a natural miracle bioactive compound against lifestyle related disorders. Lipids Health Dis 16(1):1–7
Jasna H, Nada M, Franci P, Leon O, Silvana P (2009) Wooden joints influence on durability of construction. In: Wood is good: new materials, quality and design of products. In: Proceedings of the 20th International Scientific Conference 16th October 2009, University of Zagreb, Faculty of Forestry, Zagreb, Croatia, pp. 129–136
Karanjalker GR, Ravishankar KV, Shivashankara KS, Dinesh MR, Roy TK, Sudhakar Rao DV (2018) A study on the expression of genes involved in carotenoids and anthocyanins during ripening in fruit peel of green, yellow, and red colored mango cultivars. Appl Biochem Biotechnol 184(1):140–154
Katz-Jaffe MG, McReynolds S, Gardner DK, Schoolcraft WB (2009) The role of proteomics in defining the human embryonic secretome. Mol Hum Reprod 15(5):271–277
Keathley J, de Toro-Martín J, Kearney M, Garneau V, Pilon G, Couture P, Marette A, Vohl MC, Couillard C (2022) Gene expression signatures and cardiometabolic outcomes following 8-week mango consumption in individuals with overweight/obesity. Front Nutr 9
Khandare MS (2016) Mango (*Mangifera indica* Linn) a medicinal and holy plant. J Med Plants Stud 4(4):44–46
Khurana RK, Kaur R, Lohan S, Singh KK, Singh B (2016) Mangiferin: a promising anticancer bioactive. Pharmac Patent Analyst 5(3):169–181
Kim SG, Kim SY, Park CM (2007) A membrane-associated NAC transcription factor regulates salt-responsive flowering via FLOWERING LOCUS T in Arabidopsis. Planta 226(3):647–654
Kogel KH, Voll LM, Schäfer P, Jansen C, Wu Y, Langen G, Imani J, Hofmann J, Schmiedl A, Sonnewald S, von Wettstein D (2010) Transcriptome and metabolome profiling of field-grown transgenic barley lack induced differences but show cultivar-specific variances. Proc Natl Acad Sci U S A 107(14):6198–6203
Lagaert S, Beliën T, Volckaert G (2009) Plant cell walls: Protecting the barrier from degradation by microbial enzymes. In: Seminars in Cell and Developmental Biology 2009, Vol 20, No 9, Academic Press, pp 1064–1073
Laldinchhana Lalrengpuii J, Ray S, Pachuau L (2020) Indian tropical fruits and their bioactive compounds against human diseases. Plant-Derived Bioact Chem Mode Act:455–494
Maisuthisakul P, Gordon MH (2009) Antioxidant and tyrosinase inhibitory activity of mango seed kernel by product. Food Chem 117(2):332–341
Masibo M, He Q (2008) Major mango polyphenols and their potential significance to human health. Comprehens Rev Food Sci Food Safe 7(4):309–319
Mathiazhagan M, Chidambara B, Hunashikatti LR, Ravishankar KV (2021, 1881) Genomic approaches for improvement of tropical fruits: fruit quality, shelf life and nutrient content. Genes 12(12)

Mazidi M, Katsiki N, Mikhailidis DP, Sattar N, Banach M (2019) Lower carbohydrate diets and all-cause and cause-specific mortality: a population-based cohort study and pooling of prospective studies. Eur Heart J 40(34):2870–2890

Menon B, Harinarayan CV, Raj MN, Vemuri S, Himabindu G, Afsana TK (2010) Prevalence of low dietary calcium intake in patients with epilepsy: a study from South India. Neurol India 58:209–212

Mohan V, Sandeep S, Deepa R, Shah B, Varghese C (2007a) Epidemiology of type 2 diabetes: Indian scenario. Indian J Med Res 125:217–230

Mohan V, Sudha V, Radhika G, Radha V, Rema M, Deepa R (2007b) Gene-environment interactions and the diabetes epidemic in India. Forum Nutr 60:118–126

Momeny E, Rahmati S, Ramli N (2012) Effect of microwave pretreatment on the oil yield of mango seeds for the synthesis of a cocoa butter substitute. J Food Process Technol 3(7)

Naushad SM, Radha A, Devi R (2010) Role of parental folate pathway single nucleotide polymorphisms in altering the susceptibility to neural tube defects in South India. J Perinat Med 38: 63–69

Neeha VS, Kinth P (2013) Nutrigenomics research: a review. J Food Sci Technol 50(3):415–428

OECD/FAO (2020) OECD-FAO Agricultural Outlook 2020–2029, vol 13. FAO/OECD Publishing, Rome/Paris, p 27

Papanikolaou Y, Fulgoni V (2022) Mango consumption is associated with improved fruit and nutrient intakes and a higher diet quality in US adults. Curr Develop Nutr (Suppl 1):938

Parvez GM (2016) Pharmacological activities of mango (*Mangifera Indica*): JPharmacog. Phytochemistry 5(3):1

Passos GA (2015) Transcriptomics in health and disease. Springer

Pavlou MP, Diamandis EP (2010) The cancer cell secretome: a good source for discovering biomarkers? J Proteome 73(10):1896–1906

Piccolo SR, Mecham A, Golightly NP, Johnson JL, Miller DB (2022) The ability to classify patients based on gene-expression data varies by algorithm and performance metric. PLoS Comput Biol 18(3):e1009926

Polesani M, Bortesi L, Ferrarini A, Zamboni A, Fasoli M, Zadra C, Lovato A, Pezzotti M, Delledonne M, Polverari A (2010) General and species-specific transcriptional responses to downy mildew infection in a susceptible (*Vitis vinifera*) and a resistant (*V. riparia*) grapevine species. BMC Genomics 11(1):1–6

Raj M, Sundaram KR, Paul M, Deepa AS, Kumar RK (2007) Obesity in Indian children: time trends and relationship with hypertension. Natl Med J India 20(6):288–293

Rajapaksha L, Gunathilake DC, Pathirana S, Fernando T (2021) Reducing post-harvest losses in fruits and vegetables for ensuring food security—case of Sri Lanka. MOJ Food Process Technol 9:7–16

Rao S (2001) Nutritional status of Indian population. J Biosci 26(4):481–489

Robles-Sánchez RM, Rojas-Graü MA, Odriozola-Serrano I, González-Aguilar GA, Martín-Belloso O (2009) Effect of minimal processing on bioactive compounds and antioxidant activity of fresh-cut 'Kent'mango (*Mangifera indica* L.). Postharv Bio Technol 51(3):384–390

Rubab Z, Ashraf M, Naz S, Shahid S, Rehman R (2022) Identification of SNP of VDR and VDBP gene and its Dysregulated pathway through VDR-VDBP interaction network analysis in Vitamin D deficient Infertile Females

Rymbai H, Srivastav M, Sharma RR, Patel CR, Singh AK (2013) Bio-active compounds in mango (*Mangifera indica* L.) and their roles in human health and plant defence–a review. J Hortic Sci Biotechnol 88(4):369–379

Sahu S, Das BK, Pradhan J, Mohapatra BC, Mishra BK, Sarangi N (2007) Effect of *Magnifera indica* kernel as a feed additive on immunity and resistance to Aeromonas hydrophila in Labeo rohita fingerlings. Fish Shellfish Immunol 23(1):109–118

Sandhu KS, Singh N, Lim ST (2007) A comparison of native and acid thinned normal and waxy corn starches: physicochemical, thermal, morphological and pasting properties. Food Sci Technol 40(9):1527–1536

Seidelmann SB, Claggett B, Cheng S, Henglin M, Shah A, Steffen LM, Folsom AR, Rimm EB, Willett WC, Solomon SD (2018) Dietary carbohydrate intake and mortality: a prospective cohort study and meta-analysis. Lancet Public Health 3(9):e419–e428

Shabani Z, Sayadi A (2014) The antimicrobial in vitro effects of different concentrations of some plant extracts including tamarisk, march, acetone and mango kernel. J Appl Pharmaceut Sci 4 (5):075–090

Shah KA, Patel MB, Patel RJ, Parmar PK (2010) *Mangifera indica* (mango). Pharmacog Rev 4(7):42

Slavin JL, Lloyd B (2012) Health benefits of fruits and vegetables. Adv Nutr 3(4):506–516

Sood R, Kamikubo Y, Liu P (2017) Role of *RUNX1* in hematological malignancies. J Am Soc Hematol 129(15):2070–2082

Soong YY, Barlow PJ (2006) Quantification of gallic acid and ellagic acid from longan (*Dimocarpus longan* Lour.) seed and mango (*Mangifera indica* L.) kernel and their effects on antioxidant activity. Food Chem 97(3):524–530

Stefania DS, Clodoveo ML, Cariello M, D'Amato G, Franchini C, Faienza MF, Corbo F (2021) Polyphenols and obesity prevention: critical insights on molecular regulation, bioavailability and dose in preclinical and clinical settings. Crit Rev Food Sci Nutr 61(11):1804–1826

Torres-León C, Rojas R, Contreras-Esquivel JC, Serna-Cock L, Belmares-Cerda RE, Aguilar CN (2016) Mango seed: functional and nutritional properties. Trend Food Sci Technol 55:109–117

USDA N (2015) The plantss database. National Plant Data Team, Greensboro. http://plants.usda.gov. May 2011

Wauthoz N, Balde A, Balde ES, Van Damme M, Duez P (2007) Ethnopharmacology of *Mangifera indica* L. bark and pharmacological studies of its main C-glucosylxanthone, mangiferin. Int J Biomed Pharm Sci 1(2):112–119

Zivkovic AM, German JB (2009) Metabolomics for assessment of nutritional status. Curr Opnin Clin Nutr Metabol Care 12(5):501

Genetic Enhancement of Nutraceuticals in Papaya (*Carica papaya* L.)

C. Vasugi, K. V. Ravishankar, Ajay Kumar, and K. Poornima

Contents

C. Vasugi (✉) · K. V. Ravishankar · A. Kumar · K. Poornima
ICAR-Indian Institute of Horticultural Research, Bengaluru, India
e-mail: Vasugi.C@icar.gov.in; ravishankar.kv@icar.gov.in

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_39

Abstract

Not just in India, but all around the world, papaya (*Carica papaya* L.) is considered one of the most significant fruits. It is grown in tropical and subtropical areas all over the world. Over 37 nations around the world currently cultivate it. One of the biggest issues causing economic losses in the world is the papaya ringspot virus (PRSV-P). Therefore, producing resistant cultivars with good fruit quality and yield is the primary goal of papaya improvement. The fruits are abundant in vitamins, minerals, and other chemical components that have been linked to improved health. There are bioactive components in various plant parts, viz., shoots, leaves, immature and ripe fruit, latex, roots, and seeds, and these compounds possess antioxidant, anticancer, anti-inflammatory, wound–healing, and antifungal properties. The genetic resources have potential genes for both abiotic and stresses that could be exploited in the crop improvement program combining both traditional and modern molecular approaches like genetic engineering to achieve the set target.

Keywords

Papaya · Nutritional composition · Medicinal properties · Genetic resources · Genetic diversity

1 Introduction

1.1 Agricultural Importance

One of the most significant and profitable fruit crops in the world is papaya (*Carica papaya* L.). It is native to Tropical America (Central America and Mexico) and a member of the Caricaceae family. The viral disease, papaya ringspot virus (PRSV-P), and postharvest losses are the two most serious problems affecting the papaya industry globally. The postharvest losses are quite high, ranging from 30% to 60% (Prasad and Paul 2021), while the economic yield losses due to PRSV disease are as high as 95% of expected yield (Babu and Banerjee 2018), rendering papaya orchards economically unviable.

1.2 Nutritional Composition

The papaya fruit is incredibly high in minerals, including iron, calcium, potassium, magnesium, and phosphorus, as well as carbohydrates and proteins. Additionally, it is a plentiful source of vitamins, with 100 g of fruit providing 2020 IU of vitamin A, 40 mg of vitamin B1, and 46 mg of vitamin C (Alara et al. 2020). The fruit contains about 85–90% water and total sugars from 10% to 13%. It is also abundant in carotenoids, the main ones being lycopene, β-cryptoxanthin, and β-carotene (Daagema et al. 2020).

1.3 Limitations of Conventional Breeding and Rationale for Intervention of Advanced Strategies

The papaya ringspot virus (PRSV-P) is one of the major problems in almost all papaya-growing regions of the country and worldwide. Since this illness affects most cultivars of the genus *Carica*, one long-term solution is the introgression of a gene from a wild related. The other breeding objectives are to develop dwarf/semi-dwarf stature cultivars having gynodioecious nature with good fruit quality, shelf life, and dual purpose (table and processing) having resistance to major fungal diseases and abiotic stresses like cold tolerance. The conventional breeding method takes a very long period as attaining homozygosity for the specific traits takes several generations (Cortes et al. 2017). Hence, there is a need to integrate the modern biotechnological tools like marker-assisted selection (MAS), genetic engineering, genome-assisted breeding approaches like targeting-induced local lesions In genomes (TILLING), genome-wide association studies (GWAS), genomic selection (GS), and genome editing to develop superior cultivars.

2 Nutritional Composition

2.1 Chemical Composition

Approximately 60% portion of the ripe fruit is eatable per 100 g of fresh fruit and is composed of approximately 85–90% water and 10–13% total sugars. The major sugars are 29.8 g of glucose, 21.9 g of fructose, and 48.3 g of sucrose per 100 g of edible fruit. It contains an abundant amount of vitamins and minerals but is very low in calories. It has 200 kJ/100 g of energy and 10% carbohydrates per 100 g of edible fruit (Anjana et al. 2018). The sensory properties are caused by a number of volatile chemicals, including esters, hydrocarbons, terpenes, alcohols, aldehydes, benzyl isothiocyanate, ketones, and other organic acids. The highly abundant volatile compound of papaya is linalool, which occurred in 94% in solo varieties, whereas the oxide *cis*-linalool is abundant in Taiwan varieties (Serra et al. 2016). Some of the major fatty

acids present in papaya seeds are palmitic acid (13.90–19.7%), oleic acid (70.84–79.10%), stearic acid (4.20–6.68%), linolenic acid (0.17–0.90%), gadoleic acid (0.51%), and arachidic acid (0.38–1.10%) (Dotto and Abihudi 2021). According to Saeed et al. (2014), pulp contains various nutritional components like vitamin: ascorbic acid (25.07–58.59 mg); phenolics: ferulic acid (277.49–186.63 mg), *p*-coumaric acid (229.59–135.64 mg), and caffeic acid (175.51–112.89 mg); and carotenoids: lycopene (0.36–3.40 mg), β-cryptoxanthin (0.28–1.06 mg), and β-carotene (0.23–0.50 mg). Papaya has four types of cysteine protease enzymes, viz., papain (less than 10%), glycyl endopeptidase-III and IV (23–28%), chymopapain A and B (26–30%), and caricain (14–26%) (Saeed et al. 2014).

2.2 Chemical Type and Structure

Papaya leaves contain flavonoids, saponin, tannin, alkaloids, and glycosides. The fruit is also an excellent source of carotenoids with major carotenoids lycopene (0.36–3.40 mg), β-cryptoxanthin (0.28–1.06 mg), and β-carotene (0.23–0.50 mg) (Saeed et al. 2014). The root of papaya contains benzyl isothiocyanate and glucosinolates carposide. The papaya seed oil also contains flavonoids, kaempferol, and myricetin (Adachukwu et al. 2013). Papain and chymopapain enzymes occur in the unripe fruit, and there are also some enzymes present in latex and other parts of the plant, viz., caricain, papain, chymopapain, and protease omega (Teng et al. 2019). Some enzymes are reported in papaya latex, that is, chitinase, cysteine endopeptidases, and glutaminyl cyclase (Daagema et al. 2020).

Papaya leaves are also rich in a variety of bioactive compounds such as flavonoids like quercetin, kaempferol-3-rutinoside, quercetin 3-rutinoside, and myricetin 3-rhamnoside (Nugroho et al. 2017); carotenoids like lycopene, zeaxanthin, cryptoxanthin, β-carotene, and violaxanthin; and other phytochemicals such as kaempferol, myricetin, and quercetin. The leaves are rich in phenolic compounds, viz., "kaempferol, protocatechuic acid, quercetin, 5,7-dimethoxy coumarin, caffeic acid, p-coumaric acid, and chlorogenic acid" (Canini et al. 2007).

There are many compounds present in various papaya portions, viz., pulp: linalool (Daagema et al. 2020); leaves: dehydrocarpaine I and II, carpaine, psudocarpaine, and alkaloids (Teng et al. 2019); latex: glutaminyl cyclase; shoots: quercetin, chitinases class II and III, kaempferol, and cysteine endopeptidases; and roots: cyanogenic compounds. The benzyl glucosinolate and their degradation products like benzyl isothiocyanate are present in all the tissues of papaya. According to Ghosh et al. (2017), papaya seeds have a good amount of oleic acid and the antifertility compound 1,2,3,4-tetrahydropyridin-3-yl-octanoate. The papaya pericarp, pulp, and seed also contain benzyl glucosinolate and benzyl isothiocyanate.

An important nutritional quality aspect of papaya fruit is regarded to be its pulp color. Generally, the red pulp types are rich in lycopene and the yellow pulp types are rich in carotenoids. There are some phenolic compounds also identified in papaya, viz., "ferulic acid (277.49–186.63 mg), p-coumaric acid (229.59–135.64 mg), and

Table 1 Important chemical compounds present in different parts of papaya

Plant parts	Compounds	References
Fruits	Vitamins: vitamins A, C, B1, B2, and B3 Acids: malic acid, citric acids, and amino acid Volatile compounds: linalool, benzyl isothiocyanate, *cis* and *trans* 2, 6-dimethyl-3, 6 expoxy-7-octen-2-ol Alkaloid: α; carpaine and benzyl-β-d glucoside	Adedayo et al. (2021)
Seeds	Carpaine, benzyl isothiocyanate, benzyl glucosinolate, glucotropaeolin, benzyl thiourea, caricin, and myrosin	Moses and Olanrewaju (2018)
Latex	Papain and chymopapain, glutamine cyclotransferase, chymopapain A, B, and C, peptidase A and B, and lysozymes	Ngafwan et al. (2018)
Leaves	Carpaine, pseudocarpaine, and dehydrocarpaine I and II, choline, carposide, vitamins C and E	Naureen et al. (2022)
Shoots	β-Sitosterol, glucose, fructose, sucrose, and galactose	Naureen et al. (2022)
Roots	Arposide and myrosin	Singh et al. (2021)

caffeic acid (175.51–112.89 mg)" per 100 g of fresh fruit that have antioxidant and antimicrobial properties (Alara et al. 2020) (Tables 1 and 2).

2.3 Medicinal Properties

2.3.1 Dengue Fever

Dengue fever is an infectious ailment in human beings that is caused by dengue viruses and spread by mosquitoes. Because it can occasionally induce excruciating pain in muscles and joints that feel like bones are breaking, this illness was previously known as break-bone fever. The major critical condition in case of dengue fever is thrombocytopenia, which can be alleviated by the use of papaya leaves (Sarker et al. 2021). Studies conducted in vitro revealed that papaya leaf extracts have the ability to stabilize membranes and, at lower doses, reduce heat-induced and hypotonicity-induced hemolysis of erythrocytes in both healthy and dengue-infected persons (Ranasinghe et al. 2012); thus, it may be useful in preventing platelet lysis (Fig. 1).

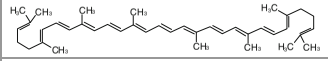
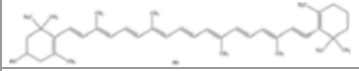
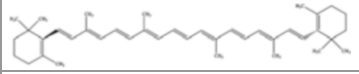
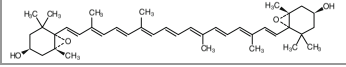
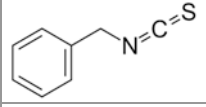
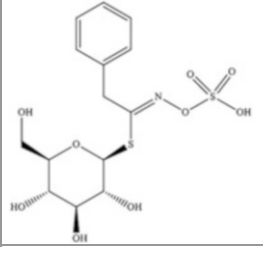
2.3.2 Anti-Inflammatory Property

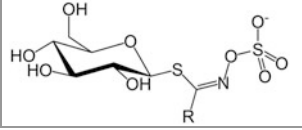
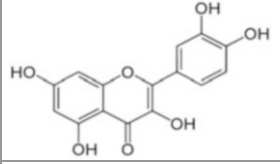
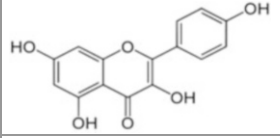
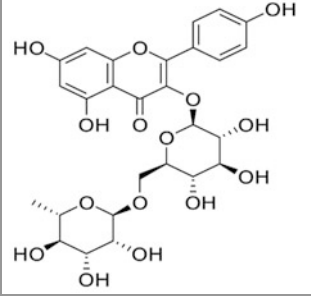
This activity found in papaya is because of the presence of enzyme cysteine proteinases. Papain was found to be safe and effective in the treatment of chronic inflammation. The anti-inflammatory property of papaya seeds was confirmed by Amazu et al. (2010)

2.3.3 Anticancer Activity

Due to the proteolytic enzymes present in papaya, which convert protein and the fibrin cancer cell wall into amino acids, the fruit possesses anticancer properties. Isothiocyanate (the degradation product of benzyl glucosinolate) is also very

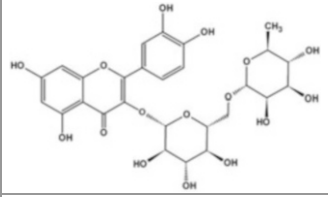
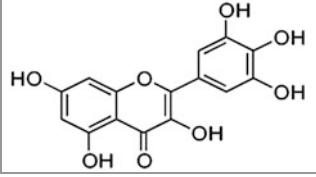
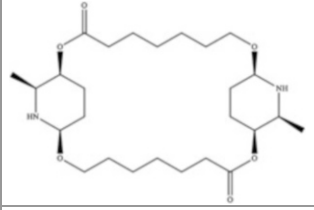
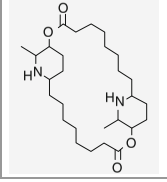
Table 2 Prominent bioactive compounds present in different parts of papaya: structure and properties

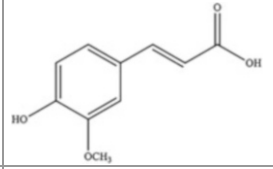
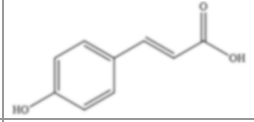
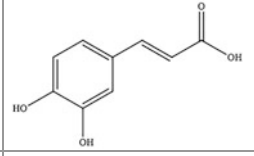
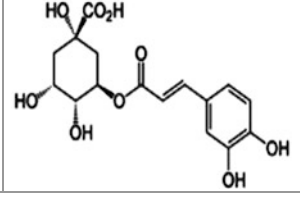
Bioactive compounds	Structure	Plant source	Properties	References
Carotenoids				
Lycopene		Fruit	Antioxidant	Adedayo et al. (2021)
β-Cryptoxanthin		Fruit	Antioxidant	Adedayo et al. (2021)
β-Carotene		Fruit	Detoxifying properties	Adedayo et al. (2021)
Violaxanthin		Fruit, leaves		Kaur et al. (2019)
Glycosides				
Benzyl isothiocyanate		Roots, pulp, seeds	Antibacterial	Pinnamaneni (2017)
Benzyl glucosinolate		Roots, pulp, seeds	Anticancer	Pinnamaneni (2017)

Glucosinolates carposide		Roots, seeds	Anticancer	Pinnamaneni (2017)
Flavonoids				
Quercetin		Shoots,	Antioxidant, antidiarrheal	Adedayo et al. (2021)
Kaempferol		Shoots, seed	Antioxidant, anticancer	Nugroho et al. (2017)
Kaempferol-3-rutinoside		Leaves	Antioxidant, anticancer	Nugroho et al. (2017)

(continued)

Table 2 (continued)

Bioactive compounds	Structure	Plant source	Properties	References
Quercetin-3-rutinoside		Leaves	Antioxidant, antidiarrheal	Nugroho et al. (2017)
Myricetin		Seeds	Antioxidant	Nugroho et al. (2017)
Alkaloids				
Carpaine		Leaves	Antioxidant	Teng et al. (2019)
Pseudocarpaine		Leaves	Antioxidant	Teng et al. (2019)

Phenolics				
Ferulic acid		Fruit, leaves	Antioxidant	Alara et al. (2020)
Oleic acid		Seeds	Antioxidant	Ghosh et al. (2017)
p-Coumaric acid		Fruit	Antioxidant	Alara et al. (2020)
Caffeic acid		Fruit, leaves	Antioxidant	Alara et al. (2020)
Chlorogenic acid		Leaves	Antioxidant, antimicrobial	Canini et al. (2007)

(continued)

Table 2 (continued)

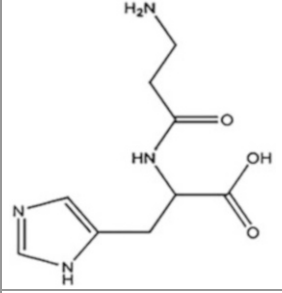
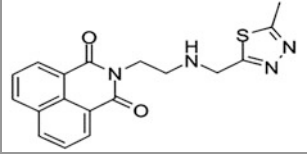
Bioactive compounds	Structure	Plant source	Properties	References
Enzymes				
Papain		Unripe fruit latex	Anti-inflammation, wound healing	Azarkan et al. (2003)
Chitinase		Papaya latex	Antifungal	Azarkan et al. (2006)

Fig. 1 Health-related beneficial role and products of papaya. (Sharma et al. 2020)

effective against various cancers, viz., colon, lung, breast, pancreas, prostate, and leukemia. The aforementioned enzymes have the ability to prevent the growth and development of cancer cells (Fauziya and Krishnamurthy 2013).

2.3.4 Antifungal Activity

The latex of papaya combined with the chemical fluconazole can inhibit the growth of *Candida albicans* (Giordani et al. 1997). Antifungal activity is attributed to latex proteins, and for complete inhibition of fungal growth the minimum quantity required is about 138 mg/dl (Dwivedi et al. 2020).

2.3.5 Wound-Healing Activity

Papaya has wound-healing property due to the presence of proteolytic enzymes. Papain is very effective in the treatment of ulcer in rats by blocking the acid secretion (Chen et al. 1981). Papain is a nonspecific cysteine proteinase that can digest the substrates of necrotic tissue (pH varied from 3.0 to 12.0). These proteolytic enzymes decrease the risk of oxidative tissue damage due to the increase in hydroxyproline content; additionally, they are capable of burn-healing activities (Chen et al. 1981) (Fig. 2 and Table 3).

2.4 Agronomic and Postharvest Techniques

As well as being consumed as fresh fruit, it can also be processed into various value-added goods while retaining the nutritional value. The raw papaya fruits can be processed into tutti-frutti, pickle, jellies, and candies, while ripe fruit can be processed into canned papaya, ready-to-serve fruit beverages, nectar, pulp, fruit bar, bars, toffee, osmotically dehydrated products, minimally processed products from cut pieces of ripe papaya, and also it can be blended with other fruit pulps to use

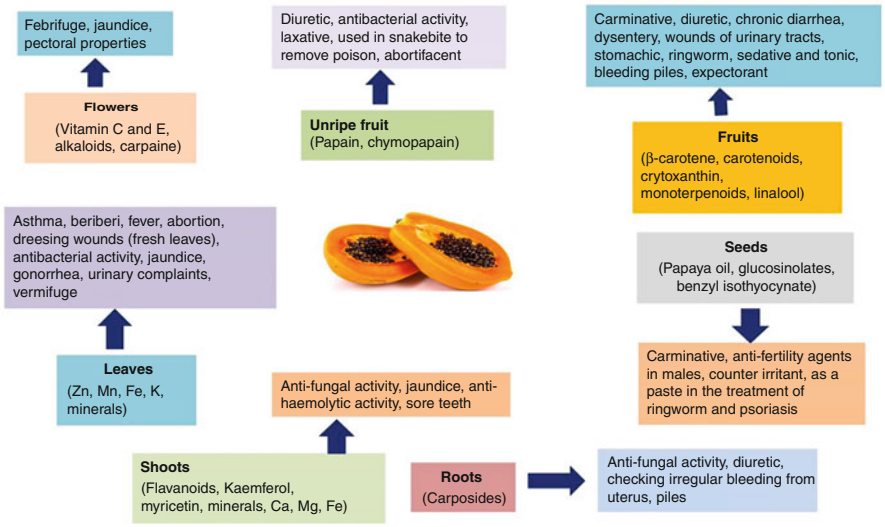


Fig. 2 Medicinal benefits of different parts of papaya

Table 3 Therapeutic benefits of various papaya portions (Anjana et al. 2018)

Part	Methods of preparation	Therapeutic uses
Peel	Use peel as a mask on your face for around 20 min	To remove skin and facial blemishes
	Peel and lemon juice apply to the scalp for 20 min	Against dandruff
	Olive, almond, and rose oil are used to stew papaya peel, and the resulting papaya oil is then used with nectar and rose water to apply onto the skin	Works as a skin tonner and skin cleanser
Fruit	Consumption of fresh fruit	Indigestion, clogging, farts, and enhanced hunger
	Apply unripe fruit on the influenced zone	Pimples, skin inflammation, and mouth ulcer
Leaves	Leaf extracts	Treating dengue fever, nervous pains, and elephantoid growth; and used to treat injuries and wounds
Root	Root infusion	Employed to treat syphilis and lessen urine concretions
Seeds	Crisp or dried seeds	Bacteriostatic, bactericidal, and fungicidal
Flowers	Flower extract	Treating jaundice
Latex	Latex of plant	Curing psoriasis, ringworm, and dyspepsia

either in juice or fruit bar. Blending of papaya to an extent of 30% produced acceptable quality of juice (Pathak et al. 2021). Papaya sauce from the extracted pulp of ripe fruits has been prepared by Ang et al. (2017). The papain is milky latex extracted from mature green fruits when lanced. It contains a protein hydrolyzing

Table 4 Utilization of different edible coatings to enhance the shelf life of papaya

Coatings	Features	References
Burkholderia cepacia B23 plus calcium and chitosan coating	The addition of $CaCl_2$ (3%) to the combination treatment raised the calcium content of the fruit (to 81%) and prolonged its shelf life	Rahman et al. (2012)
Chitosan plus peppermint essential oil	Chitosan (1%) plus peppermint essential oil (0.2%) results in less peel discoloration, good color development, and higher marketability	Picard et al. (2013)
Chitosan plus extract of propolis	Chitosan (1%) in combination with propolis ethanolic extract (5%) improved postharvest quality of fruit	Barrera et al. (2015)
Aloe vera gel	Shelf life extended up to 15 days and during storage color development improved by adding aloe vera gel coating (1.5%)	Sharmin et al. (2016)
An essential oil and carboxymethylcellulose coating (CMC)	Postharvest quality is maintained and the severity of postharvest disease is decreased when CMC is used in combination with *Lippia sidoides* essential oil	Zillo et al. (2018)

enzyme (protease or proteolytic enzyme) that has several industrial and medicinal uses. The demand for papain has significantly expanded during the last few years, and its production on commercial scale has started. The nutritional values in the processed products (tutti-frutti, pickle, jellies and candies, etc.) were estimated after 6 months of storage and found that they remain constant but the stability of dried papaya is only up to 30 days (Prasad and Paul 2021).

The use of edible coatings can improve shelf life and appearance of papaya fruit, and reduce microorganism growth and decay after harvesting, improving its postharvest quality (Prasad and Paul 2021). The quality of the fruit can be enhanced by immobilizing the fresh-cut fruit in multilayered antimicrobial coatings (chitosan and pectin) (Brasil et al. 2012). Coating of papaya fruit with chitosan (1%) plus peppermint essential oil (0.2%) results in less peel discoloration, good color development, and higher marketability (Picard et al. 2013). Fruits that are packed in HDPE and kept under evaporative cooling maintain better fresh weight and contain higher TSS and vitamin C (Table 4).

2.5 Requirement of Biotechnological Intervention

Genetic engineering is a vital approach in genetic improvement of papaya that alters one or more desirable characters in elite varieties without interfering with the prevailing traits. The development of effective gene insertion techniques to impart the desired features has helped advancements in genetic engineering. A total of 21 quantitative trait loci (QTLs) were related to seven fruit quality attributes, viz., length and breadth of fruit, sweetness and thickness of pulp, skin freckle, fruit firmness, and fruit weight (Nantawan et al. 2019). These traits could be improved through biotechnological intervention in papaya. The shelf life of papaya fruit can also be improved through co-suppression ACC oxidase genes (Lopez-Gomez et al. 2009). A bacterial color

complement test confirmed that the gene *CpCYC-b* controls papaya pulp color. Tomato chloroplast-specific *lycopene b-cyclase* and *CYC-b* is a DNA marker that is closely associated with flesh color co-localized on physical map contigs containing cDNA probes. So, papaya fruit pulp color and lycopene content can be improved through the incorporation of gene from tomato (Blas et al. 2010). The development and maturation of papaya fruit are regulated by the *CpGRFs* genes. The expression or overexpression of these genes can regulate the growth and ripening of fruit (Li et al. 2021). The markers, which are linked with many fruit quality parameters, including fruit size, pulp color, lycopene content, and carotenoid content, might facilitate marker-assisted selection and also improve quality in a short period. However, there are only a few attempts in this regard, and there is a need for more effort using modern next-generation sequencing (NGS)-based methods like QTL-seq, GWAS, etc.

3 Genetic Resources

Papaya belongs to the genus *Carica,* which is monotypic, and all the commercial varieties come under the genus *Carica*. The highly cross-pollinated nature of crop and seed multiplication is the cause of the diversity currently existing in germplasm (Chavez-Pesqueira and Nunez-Farfan 2017). Continuous research efforts made in this crop have resulted in the development of more than 25 improved cultivars. The papaya germplasm ranges from various commercial cultivars, wild types, local land races, and exotic collections. Some species of the genus *Vasconcellea* are tightly associated with *C. papaya* than others, which influence the successful use of *Vasconcellea* in papaya improvement. Commercially available cultivars are generally divided into two groups: gynodioecious (hermaphrodites and females) and dioecious (males and females). As the crop is commercially propagated through seeds, the purity is being maintained through selfing and sib mating. Several organizations around the world are involved in the conservation of germplasm in the form of seed bank and field gene banks.

However, the global germplasm of papaya has not been organized into an accessible database. The USDA site of the U.S. National Plant Germplasm System in Hilo, Hawaii, reports 153 accessions of *C. papaya* and some *Vasconcellea* spp.; EBDA, Bahia (82 accessions), EMBRAPA Mandioca e Fruticultura, Cruz das Almas, Bahia (141 accessions), and IAC Campinas, São Paulo (169 accessions); Colombia at University Nacional Medellín and CORPOICA (83 accessions) with additional accessions at other locations; and Malaysia (72 accessions) (Vincent et al. 2019). Approximately 150 papaya germplasm samples are being preserved in field gene banks in India (Indian Council of Agricultural Research – All India Coordinated Research Project or ICAR-AICRP) at a number of ICAR institutes, including the Indian Agricultural Research Institute-Research station, Pune, Maharashtra (17), TNAU-HC& RI, Coimbatore, Tamil Nadu (89), and ICAR-IIHR-National Active Germplasm Sites, Bengaluru, Karnataka (49) (http://www.aicrp.icar.gov.in) (Table 5).

Table 5 Genetic resources of different wild species as potential gene donor

Wild genetic resources	Specific traits	References
V. cundinamarcensis (syn. *V. pubescens)*	Resistant to PRSV-P, blackspot and *Erwinia* species, cold tolerant	Badillo (2000), Eeckenbrugge et al. (2014)
V. cauliflora	Resistant to PRSV-P and bacterial canker *(Erwinia papayae)*	Badillo (2000), Eeckenbrugge et al. (2014)
V. goudotiana	Resistant to *Erwinia* species, *Phytophthora,* and bacterial canker *(Erwinia papayae)*	Eeckenbrugge et al. (2014)
V. parviflora	Resistant to Paw *paw die back (Mycoplasma) and interspecies* for *V. cundinamarcensis*	Drew et al. (1998)
V. quercifolia	Resistant to PRSV-P	Badillo (2000)
V. stipulata	Resistant to PRSV-P and cold tolerance	Badillo (2000), Horovitz and Jiménez (1967)
V. candicans	Resistant to PRSV and distortion ringspot virus	Badillo (2000), Horovitz and Jiménez (1967)
V. heibornii	Resistant to PRSV and distortion ringspot virus	Horovitz and Jiménez (1967)
V. monoica	Monoecious, leaves as vegetable	Swingle (1947)
V. pentagona	Resistant to frost	Singh (1964)
V. pentandra	Cold tolerant	Hamilton and Robinson (1937)

4 Classical Genetics and Traditional Breeding

Traditional breeding and classical genetics made significant advancements and generated information for a variety of quantitative and qualitative characteristics. Superior cultivars have been developed through a variety of breeding techniques, including plant introduction, inbreeding and selection, hybridization and selection, mutation, backcross breeding, and marker-assisted selection.

4.1 Genetics of Health-Related (HR) Genes

The knowledge of various genes and their controlling mechanism, which affects economic traits, will be helpful in the selection of superior genotypes for imparting specific traits. Thus, once information about genetic pattern is generated, breeding programs are initiated with the aim of incorporating a particular desirable trait. Fruit weight has direct positive correlation with fruit size and is controlled by multiple alleles. However, the heterosis in fruit weight is demonstrated over the superior parent in the presence of overdominant gene action (Chan 2001). Total soluble solids

(TSS) and flavor are governed by a single homozygous recessive allele, which are important traits for taste and quality of fruit while there are some reports indicating that the trait TSS is determined by quantitative genes with additive effects (Rimberia et al. 2018). Pulp color is an important trait in relation to fruit quality and is controlled by a single gene. The yellow color (*R*) is predominating over red (r) and different shades of pink may be attributed to the influence of modifier genes. All red pulp (*rr*) varieties will breed true for pulp color. The green color of fruit peel is controlled by a single dominant gene (*G*) and yellow (*gg*) is governed by double-recessive gene (Aryal and Ming 2014).

Traditional breeding has been used effectively to improve qualitative traits that are directly associated with other traits, for example, carotene is linked with orange flesh color (Rimberia et al. 2018), gynodioecious is linked with fruit shape (Ming et al. 2007), and parthenocarpy is linked with seedless; these are typical examples of selection using morphological markers.

4.2 Breeding Objectives

Current breeding goals in papaya develop gynodioecious-type cultivars that are dwarf or semi-dwarf in stature and good fruit quality (high TSS, less cavity percent, good shelf life, and high pulp recovery) coupled with resistance to major fungal, viral diseases, and abiotic stress like cold tolerance.

4.3 Limitations of Conventional Breeding and Rationale for Molecular Breeding

The selection of traits, which is primarily based on morphological parameters that are affected by environmental factors, is the primary drawback of traditional breeding (Jat et al. 2021). Traditional breeding has paved the way for the development of important quantitative traits like fruit size, early maturity, and fruit yield, which is the most effective method for selection of multiple allele traits. PRSV is the major problem faced by the papaya industry, and attempts are being made to develop PRSV-resistant types through traditional breeding by incorporation of resistance from the wild relative, viz., *Vasconcellea cauliflora, V. stipulata, V. cundinamarcensis,* and *V. quercifolia*. Resistance genes from wild relatives are typically difficult to incorporate into cultivars due to cross-incompatibility between the two genera, early embryonic abortion, nonviability, lethality, and sterility of hybrid seeds. However, the postfertilization barrier to intergeneric crosses can be overcome by using growth hormones and nutrient solution to reduce embryo abortion. Sucrose at 5% was noticed to break the intergeneric barriers by promoting pollen germination (Pujar et al. 2019).

Due to their nondisruptive nature, ability to evaluate many characteristics simultaneously, and ability to eliminate trait-related environmental variation, molecular markers have the potential to overcome the constraints of conventional breeding

techniques. But in molecular-assisted breeding, the traits that are important must be identified during marker identification and a segregating population must be created in order to produce the traits. Recently, many advanced technologies of genome sequencing, linkage disequilibrium (LD) mapping, or genome-wide association studies (GWAS) have been developed for the identification of trait of interest and their important QTL regions (Zhu et al. 2008). As PRSV is the major problem in the papaya industry, molecular breeding and genome editing technologies (CRISPR Case-9) can help us to overcome the problem.

5 Genetic Diversity Analysis

The germplasm of papaya shows considerable genetic variability for various horticultural importance traits. There are various criteria that can be used in the assessment of genetic variability, including pedigree records, morphological characteristics, and molecular markers. According to Baxy (2009), comparable external morphological features that are specific to the environment and developmental stages have historically been the basis for plant taxonomy. The fact that molecular markers can be found in all plant tissues and are unaffected by environmental changes makes them ideal for plant identification. Inter-simple sequence repeats (ISSRs) are significant molecular marker that are widely used to identify various plant species and cultivars (Ahmad et al. 2010).

5.1 Phenotypic Diversity Analysis

Phenotypic characterization using descriptors related to leaf size and shape, types of flowers and inflorescence, and fruit size and shape has assisted in understanding the divergence of papaya germplasm and identifying suitable cultivars to enhance the genetic gain in papaya improvement programs (Silva et al. 2017). The flower and fruit traits are the most economically distinctive phenotypic traits used as the selection criteria for a genotype. According to the sex type of the tree, the inflorescence and flowers vary, and cultivars are basically either gynodioecious or dioecious. The female plant's inflorescence peduncle is short (2.5–6 cm long) than the male inflorescence peduncle (60–90 cm to 150 cm). The female inflorescences bear spherical- to ovoid-shaped fruits. Hermaphrodite plants have short inflorescence peduncles, are intermediate to unisexual in nature, and produce pear-shaped fruits with varying degrees of neck constriction as per the variety (Santa-Catarina et al. 2020).

As mentioned, the fruit shape corresponds to the flower type, and the flower type corresponds to plant sex. Fruit size typically ranges from 10 to 50 cm in length, depending on the shape of the fruit, which can be spherical or pear-shaped. Fruits ranged in weight from 0.35 kg to 3.6. The most divergence traits are fruit weight, number of fruits, firmness of pulp, and TSS (Santa-Catarina et al. 2020), plant height, canopy diameter, stem diameter, the average number of leaves, and an insertion height of first fruit (Silva et al. 2017).

5.2 Diversity Analysis Using DNA Markers

SSR marker libraries have been used in research, and their applications have helped in molecular-assisted selection and genetic variability evaluation. Genetic diversity analysis using 20 simple sequence repeat (SSR) markers was carried out on 164 accessions from Costa Rican and 20 known cultivars from the U.S. Department of Agriculture (USDA) (Brown et al. 2012). Genetic diversity was assessed using two SPAR (single-primer amplification reaction) techniques, that is, random amplified polymorphic DNA (RAPD, 16 primers), ISSR (12 primers), for 13 cultivars and lines, and they found a great similarity among these cultivars, Pusa delicious and Pusa Giants, Pusa Dwarf and CO-7, and Pusa Nanha and CO-2 (Sabara and Vakharia 2018). In another study, analysis of morphological traits (fruit yield traits) revealed significant diversity among the 20 genotypes of papaya, with a total of 89 polymorphic and 45 monomorphic alleles being found out of a total of 134 alleles. The diversity of morphological attributes is closely resemblance to molecular diversity analysis (Suvalaxmi et al. 2019).

5.3 Relationship with Other Cultivated Species and Wild Relatives

Papaya is a member of the Caricaceae family and is divided into six comparatively small genera such as *Carica*, *Vasconcellea*, *Horovitzia*, *Jacaratia*, *Jarilla*, and *Cylicomorpha* (Badillo 2000). The best-known and most significant species in terms of economic importance belong to the genus *Carica*, which has just one species (*Carica papaya*). The largest genus, *Vasconcellea,* has 21 species and was recently reclassified as a separate genus rather than a subgenus of the genus *Carica* (Badillo 2000). Accurate species identification can be difficult as natural interbreeding between species in the genus *Vasconcellea* is easy. The existence of hybrid *Vasconcellea* × *heilbornii and* its several variants, not all of which have been characterized, may provide an explanation. The commercially grown varieties of *V. cundinamarcensis* and *V.* × *heilbornii* cv. *babaco* possess untapped potential as sources of the proteolytic enzyme papain and genes for genetic improvement of papaya (Badillo 1993).

Many molecular studies indicated that there is a large genetic gap between *Vasconcellea* and *Carica*. Studies on intraspecific relationship among *Vasconcellea* genus showed that intraspecific *cpDNA* variation was observed in *V. microcarpa*, and *Vasconcellea* × *heilbornii* was the most diverged species (Droogenbroeck et al. 2004). Intergeneric hybridization of the *Vasconcellea* and *Carica papaya* species has shown that Arka Prabhath proved to be a good combiner with *Vasconcellea* spp. (Pujar et al. 2019).

5.4 Relationship with Geographical Distribution

According to Rimberia et al. (2018), papayas are commonly planted throughout the world's tropical and subtropical regions and are native to Mexico or South Central America. They are also said to be able to adapt to a broad spectrum of environmental

variables. Although it thrives between 36°N and 36°S, the majority of papayas are produced between 26°N and 26°S. The geographical distribution of the papaya population, both wild and cultivated, can be used to find the collection locations for their utilization and conservation. The geographical distribution of papaya is widespread in Mexico. Variable climate (latitude, altitude, temperature, and rainfall) had the greatest influence on the ecological and geographic distribution of papaya. Papaya distribution has been better understood by the identification of 16 eco-geographical areas based on climatic and edaphic geophysical characteristics (Salinas et al. 2022). The promising dispersal of papaya species in Mexico covered a total area of 114,546.5 km^2, with the coasts of Chiapas and the Gulf of Mexico possessing the highest potential. Three zones, namely the coast of Chiapas, the north part of Guerrero, and the southern region of Veracruz, are described as the highly promising spreading areas of cultivated papaya. This is done to gather genetic material and discover the many types of adaptations that exist in Venezuela's various geographical regions. The *Vasconcellea* species distribution zones have been known using the FloraMap software (Trujillo et al. 2018). DIVA-GIS software has also been used to know the distribution zones and patterns of 21 *Vasconcellea species*.

6 Molecular Mapping and QTLs for HR Genes

There have been numerous papaya genetic molecular linkage maps created during the past 70 years. Hofmeyr produced the first genetic molecular map in 1939, and it was based solely on three morphological markers, that is, flower color, types of sex, and stem color. The initial map covered a region of 41 cM and revealed that stem color is associated with flower color at a distance of 17.3 cM between the loci and distantly associated with sex type at a distance of 41 cM. This map could not be used to predict sex type due to its poor resolution. It took over 60 years for the next genetic map to be created. A second genetic molecular map was based on 62 RAPD markers and created a sex-type locus on linkage group 1. A total of 999.3 cM and 11 linkage groups (LGs) were covered by the 62 markers, with a mean distance of 19.6 cM between contiguous markers (Sondur et al. 1996). The heterozygote genotype was suppressed due to the dominant character of RAPD markers, which led to increased map distances between markers. Nevertheless, using eight more markers, the sex-type locus, *Sex1*, was finally mapped for the first time on Linkage Group 1. There was no evidence of recombination suppression in the linkage groups, including the one carrying the sex locus. The third genetic molecular map was produced based on the morphological markers such as sex type and pulp color, which comprised 1498 AFLP markers and PRSV-P coat protein marker (Ma et al. 2004). Compared to the previous linkage map, which had 19.60 cM average distance between the markers, this map represented a tremendous improvement. The flesh color of fruit served as a morphological marker and was included in the most current high-density genetic map, which had 706 SSR markers. Twelve linkage groups were created by the map (9 major and 3 minor), totaling 1068.9 cM with an average distance of 1.5 cM between the markers (Chen et al. 2007) (Table 6).

Table 6 An overview of papaya's five genetic molecular maps

Types of genetic maps and number of markers												
Stem color	Sex type	Flower color	Fruit flesh color	PRSV coat protein	RAPD	AFLP	SSR	No. of loci	Total cM mapped	No. of LG	Average distance (cM)	References
1	1	1	–	–	–	–	–	3	41	n/a	20.5	Hofmeyr (1939)
–	1	–	–	–	61	–	–	62	999	11	19.6	Sondur et al. (1996)
–	1	–	1	1	–	1498	–	1501	3294	12	2.2	Ma et al. (2004)
–	–	–	1	–	–	–	706	707	1069	12	1.5	Chen et al. (2007)
–	–	–	1	–	–	277	712	990	945	14	1.5	Blas et al. (2009)

The results of the QTL study are anticipated to show that the alleles for fruit size may control cell division and growth, while the allele for fruit shape may influence floral development (Blas et al. 2012). Genetic linkage map was constructed in which 21 QTLs were observed for seven fruit quality characteristics, viz., length and breadth of fruit, thickness and sweetness of pulp, fruit weight, skin freckle, and fruit firmness. The QTLs were unable to forecast the fruit shape and size in individuals of the F1 and F2 progenies from the intraspecific hybrids. However, there are no studies related to fruit qualities and nutrient content.

7 Marker-Assisted Breeding

7.1 Germplasm Characterization

Papaya cultivars cannot often be identified by using morphological markers till the fruit production (Nishimwe et al. 2019). The papaya cultivars and accessions were characterized by using the 65 agronomic and morphological parameters. The morphological characterization of 60 Kenyan papaya germplasm has been carried out, and results revealed that some traits showed great variability, viz., tree habit, flower color, fruit diameter, fruit shape, and leaf size (Nishimwe et al. 2019). Molecular germplasm characterization of papaya has been done using different molecular markers AFLP (Oliveira et al. 2011), SSR (Pirovani et al. 2021), and ISSR (Hassan et al. 2022).

Using ten polymorphic simple sequence repeats, 31 papaya genotypes from Spain, Brazil, Ecuador, China, Taiwan, India, and multiple spots in Bangladesh were genotyped. The P3K1024CC and P6K900CC markers had the largest numbers of alleles, great gene diversity, and polymorphism information richness (Hasibuzzaman et al. 2020). Based on the microsatellite markers, molecular characterization was done using SSR markers of 23 elite lines of papaya (Pirovani et al. 2021). In a study involving various genera of Caricaceae, the cultivated type of *Carica* is closely associated with the genus *Jarilla* and *Horovitzia* but diverged from the *Vasconcellea* species.

7.2 Marker-Assisted Gene Introgression

With a density of one per 0.7 kb, microsatellites are the most prevalent kind of tandem repetition in the papaya genome. However, it merely makes up 0.19% of the papaya's whole genome (Wang et al. 2008). Using WGG (whole-genome genotyping) and the Illumina Miseq platform, a total of 28,451 SNPs with a Transition/Transversion (Ts/Tv) ratio of 2.45 and 1982 small InDels (insertions/deletions) were recognized in order to forecast the effects of the identified variants and produce a list of ripening-related genes (RRGs) with associated variants. A total of 106 RRGs were identified to be linked with 460 variations; these variations might be converted into PCR markers to facilitate the genetic improvement of papaya through marker-assisted selection (MAS) for particular traits (Bohry et al. 2021).

Microsatellite markers were used to examine the first backcross generation of papaya (BC1S) to assess the parental genomic ratio, degree of homozygosity, and gene or allele transfer that imparts the golden fruit color traits (Pinto et al. 2013). Markers are linked to HR traits such as pulp color (marker CPFC1), sex type (CPM1815Y52 marker), lycopene content, ripening-related genes, and other fruit quality traits (Bohry et al. 2021). The MAS helped to improve the health-related traits, viz., lycopene, pulp color, sweetness, and carotenoid content.

8 Map-Based Cloning

The genome size of papaya is quite small (372 Mbp) because of this, which makes it an ideal fruit crop for genetic study (Sharma et al. 2016). Information on map-based cloning in relation to the nutritional qualities of papaya is lacking. High-density genetic maps are necessary to isolate and clone desirable genes, deconstruct genomes, and MAS. The enzyme that promotes the conversion of lycopene to beta carotene, papaya *lycopene-b-cyclase* or *CpLCY-b,* was cloned by Blas et al. (2010). However, there was no difference in the expression of *CpLCY-b* between yellow and red pulp fruit. A chloroplast-specific *CpLCY-b* was expressed seven times more in leaves than in fruit. The pulp color locus was located at the end of LG-5 on a high-density genetic molecular map employing SSR markers, with the adjacent marker being 13 cM distant.

The crystal structure of the complex between papain and cystatin B served as the basis for the design and synthesis of the tripodal synthetic papain inhibitors. The tripodal molecular construct was created by synthesizing it using the tri-azacyclophane (TAC) scaffold, simulating the discontinuous cystatin B epitope that is related in the interaction with papain. A b-hairpin loop, an N-terminal peptide segment, as well as C-terminal peptide segment are the three distinct peptide segments that help cystatin B bind to papain. When *CysTACtins* 5, 7, and 9 were examined for their ability to inhibit papain, *CysTACtin 9* demonstrated outstanding papain inhibition with a Ki of 12 nm, which is similar to cystatin B's inhibitory efficacy (Ki = 0.12 nm) (Zoelen et al. 2007).

9 Recent Concepts and Strategies

Introduction of germplasm is vital in the growth and development of papaya. Various breeding approaches were followed for varietal improvement. Selection and sib mating has been the most successful method. Also, hybridization and selection to achieve homozygosity for the traits have given considerable success. Now the focus is on the development of PRSV disease-resistant cultivars using the wild genus *Vasconcellea* as a source of resistance. Genetic engineering will assist in the speedy development of novel variety by facilitating the direct introduction of genes into elite lines as conventional breeding requires a lot of time and effort (Alvarez et al. 2021). This is done to create new cultivars with better nutritional qualities by either decreasing the level of a particular antinutrient or increasing the concentration of essential

ingredients like vitamins and carotenoids (Sabbadini et al. 2021). Recent developments in molecular methods paved the way for the discovery of potential genes that regulate specific classes of nutrient components in papaya, viz., *"CpCYC-B, CpLCY-B, CpPDS2, CpZDS, CpLCY-E, CpCHY- B"* are for carotenoids (Sabbadini et al. 2021); CpLIS1 and CpP450-2 for linalool and linalool oxide (Zhou et al. 2021).

9.1 Genome Editing

Genome editing is one of the most significant recent advancements in genetic improvement of crop, and methods based on the adaptable CRISPR/Cas9 technology have been developed for a variety of features, including improvement in yield, improving resistant or tolerant to abiotic, abiotic stresses and pests, and alteration of genetics of plants for quality fruit. The TILLING technology is used for improved storage life with better fruit quality of papaya (Gauffier et al. 2016). The *V. pubescens* chloroplast genome was constructed and sequenced using Oxford Nanopore Technology. According to Lin et al. (2017), the genome size is 158,712 bp in length, which is less than the chloroplast genome of *C. papaya* (160,100 bp). These two structural haplotypes, *LSC IRa SSCrc IRb* and *LSC IRa SSC_ IRb,* were observed in the chloroplast genomes of *V. pubescens* and *C. papaya*. The chloroplast genome of *C. papaya* has 46 RNA editing locations with an approximate RNA editing effectiveness of 63%. The chloroplast genome, together with the nuclear and mitochondrial genomes, is an essential part of the plant genomes in a species, facilitating adaptation, diversification, and the emergence of plant lineages. Based on the above finding, *V. pubescens* can contribute to crop improvement in *C. papaya*. To create a visually scoreable albino phenotype in altered tissue, Brewer and Chambers (2022) used CRISPR/Cas9 to target the putative "*C. papaya* L. *phytoene desaturase* (*CpPDS*)" gene. A total of 73 plant lines were successfully transformed using the CRISPR construct *pAC0025*, which targets the *CpPDS* gene. 59 of them (81%) were entirely albinos. Ten *pAC0025* (*CpPDS* construct) lines, one untransformed control line, and one transformed line with the negative regulation construct *pAC0026* were genotyped for the targeted area in *CpPDS* (no *gRNA* construct). All three *gRNA* target locations showed mutations. Overall, high frequency of mutations found at *CpPDS gRNA* target locations and the high percentage of recovered albino plants point to a successful genome editing procedure that may be utilized to enhance papaya at the genetic level. However, there is no information available on work related to improvement of fruit quality.

9.2 Nanotechnology

Nanotechnology has been one of the most dynamic and emerging areas of science. In comparison to other physicochemical approaches, using the plant material fabrication of nanoparticles (NPs) results in more well-specified sizes and morphologies. Because plant materials operate as capping and reducing agents, which frequently aid in minimizing NPs' agglomeration, plant extracts utilized for the creation of

nanoparticles are more favorable than chemical synthesis. ZnO NPs have been synthesized using *Carica papaya* L leaf extract and characterized using UV-V in spectrum, X-ray diffraction, etc. The seeds of chickpea were treated with various concentrations of ZnO NPs, and the seed germination, shoot length, root length, and antioxidant enzyme were studied (Dulta et al. 2021). Green chemistry-based strategies for creating NPs have become a reality in recent years. Using papaya peel bio waste, the copper oxide NPs (CuO NPs) have been developed. These NPs were used as a photocatalyst to help palm oil mill effluent degrade when exposed to UV light (Phang et al. 2021). NPs based on extract derived from different parts of papaya have been extensively put into use as a green technology.

10 Genetic Engineering

The very first prerequisite for genetic engineering is the availability of plant regeneration protocols. It was effective to use premature zygotic embryos and juvenile seedling tissues in the papaya regeneration process developed by Fitch (2005). In order to obtain nutritionally rich papaya, it has to be grown in a disease-free condition. Two PRSV (papaya ringspot virus) resistant cultivars 'SunUp' and 'Rainbow' have been developed using a concept called parasite-derived resistance wherein the coat protein-mediated gene silencing mechanism is used (Gaskill 2001). Another significant problem with papaya is fast fruit ripening, and this problem has been addressed by downregulating "*1-aminocyclopropane-1-carboxylate (ACC)* synthase" enzyme involved in the biosynthesis of ethylene. In papaya, the ACC synthase enzyme, which is an important precursor in the biosynthesis of ethylene, has been downregulated to delayed fruit ripening. The genetics of papayas have undergone yet another modification to the ethylene perception pathway to postpone fruit ripening (Fitch 2005). A low-temperature stress tolerance was brought about by transgenics overexpressing *CBF1 gene*. This gene is responsible for the expression of *COR genes* that give freezing tolerance to papaya plants.

11 Role of Bioinformatics as a Tool

The multidisciplinary field of bioinformatics combines "biology, computer science, mathematics, and statistics." It is used to extract, analyze, integrate, and present the biological data generated by omics platform technologies. There are several biological databases and bioinformatics tools available to other researchers working in related subjects.

11.1 Gene, Genome, and Comparative Genome Databases

The genome of the papaya has been sequenced. 90% of the euchromatic sections and 75% (277.4 Mb) of the genome (372 Mb) were represented by the sequences. 16,362

unigenes were included in the assembled EST unigene set, and the assembled entire shotgun sequences matched 92.1% of them. The transcribed sequences accounted for 48% of the genic area and 3.6% (13.4 Mb) of the whole genome in the WGS (Ming et al. 2008). Unfortunately, there has not been any progress in the genome data sequences. With the aid of this database, we can locate promising genes related to fruit quality, growth, and development.

The papaya genome sequence can be searched through these databases: "NCBI, PlantGDB, Phytozome, and PLAZA." To examine the similarity and homology of the sequences, two initiatives have been introduced: HMMER and BLAST. "InterProScan, SMART, and Pfam" tools were developed to identify the gene family, domain association, and motif sequence, respectively. PLACE and PlantCARE tools are designed to identify the *cis*-regulatory elements and promoter of the desirable genes (Alok et al. 2019).

The papaya genome sequence can be retrieved from open-access databases to start a comparative genomics investigation. The abovementioned databases are also used for the comparative genomics analysis. The gene sequences are also aligned using the multiple sequence alignment programs ClustalX and MAFFT. The MEGA tool is often used to create phylogenetic trees (Kumar et al. 2018).

11.2 Gene Expression Databases

These are the EBI ArrayExpress (EBI AE) and the NCBI Gene Expression Omnibus in an MIAME-compliant way. Contrary to the International Nucleotide Sequence Database, these two gene expression databases have not been exchanging information. As of 2017, AE no longer imports data from GEO. Furthermore, the DNA DataBank of Japan (DDBJ) recently started a similar repository called the Genomic Expression Archive (GEA). The proteomics data of MS origin can be submitted to a public database like the "ProteomeXchange Consortium" in order to promote reproducibility of proteomics data (Abidin et al. 2021) (Fig. 3).

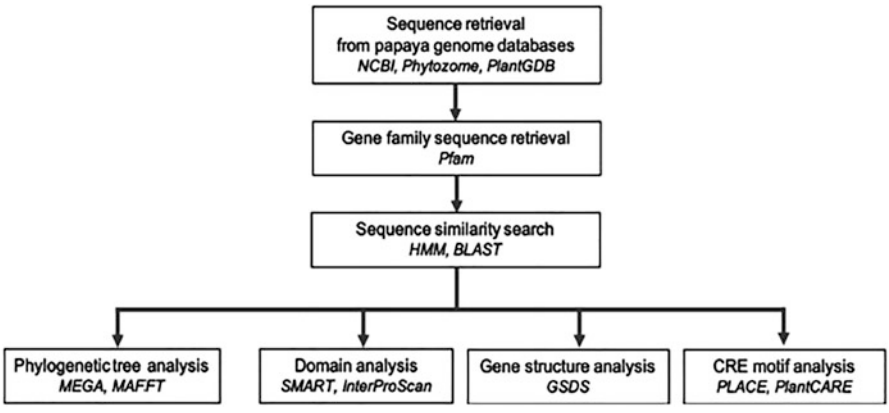


Fig. 3 Schematic approaches for comparative genomics analysis of papaya. (Abidin et al. 2021)

11.3 Protein or Metabolome Databases

There are some protein databases for papaya such as UniProt comprises UniProtKB for protein knowledge, UniRef for sequence cluster, UniParc for sequence archive, and another section is proteomes; NCBI; PlantGDB comprises CpGDB for papaya genomes; and RCSB PDB for the structure of a protein (1D–3D view) along with electron density (Abidin et al. 2021).

12 Social, Political, and Regulatory Issues

The nutritional security of expanding global population can greatly benefit from the health advantages of papaya plants. The major issues faced by papaya growers involve unfavorable weather, pest and disease incidence, nonavailability of disease-resistant good commercial cultivars, lack of processing and storage facilities, and huge postharvest losses. Awareness toward its nutritional and health benefit including the market potential of papaya is needed to boost its cultivation besides its use as a nutritious alternative. More importantly, developing postharvest techniques and value-added products that retain the phytochemicals and antioxidants present in the fresh fruit could be a boon to the industry and consumer market. Further, regulatory issues related to its biomedical application, nanotechnology, and alternative therapies need to be streamlined for its safe usage.

13 Future Prospects

Due to the significant genetic diversity that exists in papaya germplasm around the world and its importance as a fruit crop, it is possible to use this germplasm to develop a variety of improved traits. Though the crop has very good pharmacological value, viz., antioxidant, anticancer, anti-inflammatory, wound healing, antifungal and dengue cure, sufficient information is not available on these aspects and needs to be generated. There is a need to improve whole-genome sequencing data and develop a papaya genome database for easy access to genomic information. Integration of conventional breeding approach with recent biotechnological intervention like marker-assisted selection (MAS), marker-assisted introgression (MAI), association studies, genomic selection, genome editing, and transformation for health-related traits in addition to disease resistance needs to be intensified. Papaya cultivation for processing and papain production has become a profitable venture in a few countries that can be extended to other papaya-growing regions also. Very limited literature is available on map-based cloning and genomic libraries pertaining to health-related traits at present that needs more focus. Adoption of efficient production technologies like water budgeting and carbon sequestration, and integrated crop protection techniques for major fungal and viral diseases are some of the pressing needs that should be looked into and studied.

14 Conclusions

Papaya is one of the most important fruit crops of India as well as other tropical countries all over the world. The fruits are rich source of vitamins, minerals, and other bioactive compounds, which are linked to improve the health-related traits. It has wide genetic diversity, which is contributing toward the development of improved varieties with desirable horticultural traits with improved fruit quality, including resistance to viral disease like papaya ring-spot virus. As the traditional breeding has some limitations, use of molecular approaches like MAS, genome-wide association studies (GWAS), transgenic approaches, and genome editing methods like CRISPR/Cas can help to speed up the results in improving various agronomic and health-related traits in addition to resistance or tolerance to biotic and abiotic stresses.

References

Abidin RA, Ruhaizat-Ooi IH, Harun S (2021) A review of omics technologies and bioinformatics to accelerate improvement of papaya traits. Agronomy 11(7):1356

Adachukwu IP, Ogbonna AO, Eze FU (2013) Phytochemical analysis of paw-paw (*Carica papaya*) leaves. Intl J Life Sci Biotechnol Pharma Res 2(3):347–351

Adedayo BC, Oyeleye SI, Okeke BM, Oboh G (2021) Anti-cholinesterase and antioxidant properties of alkaloid and phenolic-rich extracts from pawpaw (*Carica papaya*) leaf: a comparative study. Flav Frag J 36(1):47–54

Ahmad I, Bhagat S, Sharma TVRS, Kumar K, Simachalam P, Srivastava RC (2010) ISSR and RAPD markers based the DNA fingerprinting and diversity assessment of *Annona* spp in South Andamans. Indian J Hortic 67(2):147–151

Alara OR, Abdurahman NH, Alara JA (2020) Carica papaya: comprehensive overview of the nutritional values, phytochemicals and pharmacological activities. Adv Trad Med 22:1–31

Alok A, Tiwari S, Tuli R (2019) A bidirectional promoter from Papaya leaf crumple virus functions in both monocot and dicot plants. Physiol Mol Plant Pathol 108:101423

Alvarez D, Cerda-Bennasser P, Stowe E, Ramirez-Torres F, Capell T, Dhingra A, Christou P (2021) Fruit crops in the era of genome editing: closing the regulatory gap. Plant Cell Rep 40(6):915–930

Amazu LU, Azikiwe CCA, Njoku CJ, Osuala FN, Nwosu PJC, Ajugwo AO, Enye JC (2010) Antiinflammatory activity of the methanolic extract of the seeds of *Carica papaya* in experimental animals. Asian Pac J Trop Med 3(11):884–886

Ang EM, Abdullah N, Muhammad N, Lizardo RCM (2017) Effect of formulation with papaya sauce as fat replacer on butter cake texture. J Sci Technol 9(4):56–63

Anjana GV, Priya D, Srimathi R, Shantha KB (2018) A review on medical advantages and chemical constituents of *Carica papaya* Linn. Asian J Pharm Clin Res 11(9):53–57

Aryal R, Ming R (2014) Cloning major genes controlling fruit flesh color in papaya. In: Ming R (ed) Genetics and genomics of papaya. Springer, New York, pp 341–353

Azarkan M, El Moussaoui A, Van Wuytswinkel D, Dehon G, Looze Y (2003) Fractionation and purification of the enzymes stored in the latex of *Carica papaya*. J Chromatogr B 790(1–2): 229–238

Azarkan M, Dibiani R, Goormaghtigh E, Raussens V, Baeyens-Volant D (2006) The papaya Kunitz-type trypsin inhibitor is a highly stable β-sheet glycoprotein. Biochim Biophys Acta Proteins Proteom 1764(6):1063–1072

Babu KS, Banerjee A (2018) Biological and molecular evidence of papaya ring spot virus pathotype P from mid-hills of Meghalaya, India. Indian Phytopathol 71(4):611–620

Badillo VM (1993) Caricaceae. Segundo esquema. Rev Fac Agron Univ Centr Venezuela 43:1–111
Badillo VM (2000) *Carica* L. Vs. *Vasconcellea* St. Hil. (Caricaceae) con La rehabilitation de else ultimo. Ernstia 10:74–99
Barrera E, Gil J, Restrepo A, Mosquera K, Durango D (2015) A coating of chitosan and propolis extract for the postharvest treatment of papaya (*Carica papaya* L. cv. Hawaiiana). Rev Fac Nacional de Agrono Medellín 68(2):7667–7678
Baxy JN (2009) Morphological characterization of four locally available Papaya Cultivars. B. Sc. (Hons) Dissertation, UOM Library, University of Mauritius
Blas AL, Yu Q, Chen C, Veatch O, Moore PH, Paull RE, Ming R (2009) Enrichment of a papaya high-density genetic map with AFLP markers. Genome 52(8):716–725
Blas AL, Ming R, Liu Z, Veatch OJ, Paull RE, Moore PH, Yu Q (2010) Cloning of the papaya chromoplast-specific lycopene β-cyclase, CpCYC-b, controlling fruit flesh color reveals conserved microsynteny and a recombination hot spot. Plant Physiol 152(4):2013–2022
Blas AL, Yu Q, Veatch OJ, Paull RE, Moore PH, Ming R (2012) Genetic mapping of quantitative trait loci controlling fruit size and shape in papaya. Mol Breed 29(2):457–466
Bohry D, Ramos HCC, Dos Santos PHD, Boechat MSB, Arêdes FAS, Pirovani AAV, Pereira MG (2021) Discovery of SNPs and InDels in papaya genotypes and its potential for marker assisted selection of fruit quality traits. Sci Rep 11(1):1–8
Brasil IM, Gomes C, Puerta-Gomez A, Castell-Perez ME, Moreira RG (2012) Polysaccharide-based multilayered antimicrobial edible coating enhances quality of fresh-cut papaya. LWT Food Sci Technol 47(1):39–45
Brewer SE, Chambers AH (2022) CRISPR/Cas9-mediated genome editing of phytoene desaturase in *Carica papaya* L. J Hortic Sci Biotechnol 97:1–13
Brown JE, Bauman JM, Lawrie JF, Rocha OJ, Moore RC (2012) The structure of morphological and genetic diversity in natural populations of *Carica papaya* (Caricaceae) in Costa Rica. Biotropica 44(2):179–188
Canini A, Alesiani D, D'Arcangelo G, Tagliatesta P (2007) Gas chromatography-mass spectrometry analysis of phenolic compounds from *Carica papaya* L. leaf. J Food Compos Anal 20:584–590
Chan YK (2001) Breeding papaya (*Carica papaya* L.). In: Mohan Jain S, Priyadarshan PM (eds) Breeding plantation tree crops: tropical species. Springer, New York, pp 121–159
Chávez-Pesqueira M, Núñez-Farfán J (2017) Domestication and genetics of papaya: a review. Front Ecol Evol 5:155
Chen CF, Chen SM, Chow SY, Han PW (1981) Protective effects of *Carica papaya* Linn on the exogenous gastric ulcer in rats. Am J Chin Med 9(3):205–212
Chen C, Yu Q, Hou S, Li Y, Eustice M, Skelton RL, Veatch O, Herdes RE, Diebold L, Saw J, Feng Y (2007) Construction of a sequence-tagged high-density genetic map of papaya for comparative structural and evolutionary genomics in brassicales. Genetics 177(4):2481–2491
Cortes DFM, Catarina RS, Barros GBDA, Arêdes FAS, Silveira SFD, Ferreguetti GA, Ramos HCC, Viana AP, Pereira MG (2017) Model-assisted phenotyping by digital images in papaya breeding program. Sci Agric 74:294–302
Daagema AA, Orafa PN, Igbua FZ (2020) Nutritional potentials and uses of pawpaw (*Carica papaya*): a review. Eur J Nutr Food Saf 12:52–66
Dotto JM, Abihudi SA (2021) Nutraceutical value of *Carica papaya*: a review. Sci Afr 13:e00933
Drew RA, Magdalita PM, Godwin ID, Adkins SW (1998) An efficient interspecific hybridisation protocol for *Carica papaya* L.× *C. cauliflora* Jacq. Aust J Exp Agric 38(5):523–530
Droogenbroeck B, Kyndt T, Maertens I, Romeijn-Peeters E, Scheldeman X, Romero-Motochi JP, Van Damme P, Goetghebeur P, Gheysen G (2004) Phylogenetic analysis of the highland papayas (*Vasconcellea*) and allied genera (Caricaceae) using PCR-RFLP. Theor Appl Genet 108(8): 1473–1486
Dulta K, Koşarsoy Ağçeli G, Chauhan P, Jasrotia R, Chauhan PK (2021) Ecofriendly synthesis of zinc oxide nanoparticles by Carica papaya leaf extract and their applications. J of Cluster Sci 1–15

Dwivedi MK, Sonter S, Mishra S, Patel DK, Singh PK (2020) Antioxidant, antibacterial activity, and phytochemical characterization of *Carica papaya* flowers. Beni Suef Univ J Basic Appl Sci 9(1):1–11

Eeckenbrugge GC, Drew R, Kyndt T, Scheldeman X (2014) *Vasconcellea* for papaya improvement. In: Ming R (ed) Genetics and genomics of papaya. Springer, New York, pp 47–79

Fauziya S, Krishnamurthy R (2013) Papaya (*Carica papaya*): source material for anticancer. CIB Tech J Pharm Sci 2(1):25–34

Fitch M (2005) Carica papaya papaya. In: Litz RE (ed) Biotechnology of fruit and nut crops. CABI Publishing Wallingford, Oxon, pp 174–207

Gaskill DA (2001) Relationship Between Transgene mRNA and Protein Expression Levels, Plant Development, and Resistance to an Isolate of the Papaya Ringspot Virus in Transgenic Carica papaya. Uni of Hawai'i at Manoa

Gauffier C, Lebaron C, Moretti A, Constant C, Moquet F, Bonnet G, Caranta C, Gallois JL (2016) A TILLING approach to generate broad-spectrum resistance to potyviruses in tomato is hampered by eIF4E gene redundancy. The Plant J 85(6):717–729

Ghosh S, Saha M, Bandyopadhyay PK, Jana M (2017) Extraction, isolation and characterization of bioactive compounds from chloroform extract of *Carica papaya* seed and it's in vivo antibacterial potentiality in Channa punctatus against Klebsiella PKBSG14. Microb Pathog 111: 508–518

Giordani R, Gachon C, Moulin-Traffort J, Regli P (1997) A synergistic effect of Carica papaya latex sap and fluconazole on Candida albicans growth. Mycoses 40(11–12):429–437

Hamilton TP, Robinson TR (1937) Improvement of subtropical fruit crops (No. 1589). US Department of Agriculture, Washington, DC

Hasibuzzaman ASM, Islam AKM, Miah M, Hasan M (2020) Phylogeographic diversity and population structure of *Carica papaya* L. revealed through nuclear microsatellites. Braz J Bot 43(1):147–154

Hassan KH, Abd-Elkawy MMA, Sami R, Al Masoudi LM, Alyamani A, Aloufi S, Alharthi S, Zewail RM, El-Gioushy SF (2022) Morphologic and molecular evaluation of papaya strains (*Carica papaya* L.) under Mediterranean conditions. J Biobaased Mater Bioenergy 16(2):231–239

Hofmeyr JDJ (1939) Sex-linked inheritance in *Carica papaya* L. S Afr J Sci 36(12):283–285

Horovitz S, Jiménez H (1967) Cruzamientos interespecíficos e intergenéricos en Caricaceas y sus implicaciones fitotécnicas. Agron Trop (Maracay) 17:323–343

Indian Council of Agricultural Research (2017) AICRP on fruits (AICRP-Fruits). https://aicrp.icar.gov.in/fruits/achievements/plant-genetic-resources/

Jat GS, Behera TK, Lata S, Kumar S (2021) Classical genetics and traditional breeding in cucumber (*Cucumis sativus* L.). In: Wang H (ed) Cucumber economic values and its cultivation and breeding. IntechOpen, London, pp 201–215

Kaur M, Talniya NC, Sahrawat S, Kumar A, Stashenko EE (2019) Ethnomedicinal uses, phytochemistry and pharmacology of Carica papaya plant: A compendious review. Mini-Rev Org Chem 16(5):463–480

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35(6):1547

Li M, Ren L, Zou Z, Hu W, Xiao S, Yang X, Ding Z, Yan Y, Tie W, Yang J, Guo A (2021) Identification and expression analyses of the special 14–3-3 gene family in papaya and its involvement in fruit development, ripening, and abiotic stress responses. Biochem Genet 59(6): 1599–1616

Lin Q, Singh R, Yu Q (2017, October) Isolation and characterization of translation initiation factor 4E in *Carica papaya* and *Vasconcellea* species. In: V International Symposium on Papaya, October 24–27, 2017 Merida Yucatan, Mexico, pp 39–44

Lopez-Gomez R, Cabrera-Ponce JL, Saucedo-Arias LJ, Carreto-Montoya L, Villanueva-Arce R, Díaz-Perez JC, Gómez-Lim MA, Herrera-Estrella L (2009) Ripening in papaya fruit is altered by ACC oxidase cosuppression. Transgenic Res 18(1):89–97

Ma H, Moore PH, Liu Z, Kim MS, Yu Q, Fitch MM, Sekioka T, Paterson AH, Ming R (2004) High-density linkage mapping revealed suppression of recombination at the sex determination locus in papaya. Genetics 166(1):419–436

Ming R, Yu Q, Moore PH (2007) Sex determination in papaya. Semin Cell Dev Biol 18(3):401–408

Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL, Salzberg SL (2008) The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). Nature 452(7190):991–996

Moses MO, Olanrewaju JO (2018) Proximate and selected mineral composition of ripe pawpaw (Carica papaya) seeds and skin. J Sci Innov Res 7(3):75–77

Nantawan U, Kanchana-Udomkan C, Bar I, Ford R (2019) Linkage mapping and quantitative trait loci analysis of sweetness and other fruit quality traits in papaya. BMC Plant Biol 19(1):1–11

Naureen I, Saleem A, Mobeen A, Manj NW, Sultani HMI, Saeed S, Raheem I, Adnan M (2022) Therapeutic potential of *Carica papaya* leaves in dengue associated thrombocytopenia. Sch Bull 8(3):79–83

Ngafwan N, Wardana ING, Wijayanti W, Siswanto E (2018) The role of NaOH and papaya latex bio-activator during production of carbon nanoparticle from rice husks. Adv Nat Sci Nanosci Nanotechnol 9(4):045011

Nishimwe G, Okoth E, Rimberia F (2019) Evaluation of physicochemical, nutritional and sensory quality characteristics of new papaya hybrids fruits developed in JKUAT. J Food Res 8(1):12–15

Nugroho A, Heryani H, Choi JS, Park HJ (2017) Identification and quantification of flavonoids in *Carica papaya* leaf and peroxynitrite-scavenging activity. Asian Pac J Trop Biomed 7(3):208–213

Oliveira EJ, dos Santos Silva A, de Carvalho FM, Dos Santos LF, Costa JL, de Oliveira Amorim VB, Dantas JLL (2011) Polymorphic microsatellite marker set for *Carica papaya* L. and its use in molecular-assisted selection. Euphytica 173(2):279–287

Pathak N, Dubey S, Verma S (2021) Sensory evaluation and nutritional composition of developed papaya-jam and papaya candy. J Pharmacog Phytochem 10(1):207–210

Phang YK, Aminuzzaman M, Akhtaruzzaman M, Muhammad G, Ogawa S, Watanabe A, Tey LH (2021) Green synthesis and characterization of CuO nanoparticles derived from papaya peel extract for the photocatalytic degradation of palm oil mill effluent (POME). Sustainability 13(2):796

Picard I, Hollingsworth RG, Wall M, Nishijima K, Salmieri S, Vu KD, Lacroix M (2013) Effects of chitosan-based coatings containing peppermint essential oil on the quality of post-harvest papaya fruit. Int J Postharvest Technol Innov 3(2):178–189

Pinnamaneni R (2017) Nutritional and medicinal value of papaya (*Carica papaya* Linn.). World J Pharm Pharmaceut Sci 6(8):2559–2578

Pinto FO, Pereira MG, Luz LN, Cardozo DL, Ramos HCC, Macedo CMP (2013) Use of microsatellite markers in molecular analysis of segregating populations of papaya (*Carica papaya* L.) derived from backcrossing. Genet Mol Res 12(3):2248–2259

Pirovani AAV, Ramos HCC, Duarte RP, Miranda DP, Bohry D, Santana JGS, Santos PHAD, Pereira MG (2021) Molecular characterization of elite lines of papaya (*Carica papaya* L.) via SSR markers. Funct Plant Breed J 3(1):49–58

Prasad K, Paul JR (2021) Postharvest Losses of Papaya and Practice for Management. Food Sci Rep 2(1):7

Pujar DU, Vasugi C, Adiga D, Honnabyraiah MK, Vageeshbabu HS, Jayappa J (2019) Study on cross compatibility of intergeneric hybridization between *Carica* with *Vasconcellea* species, vol 8, p 1514

Rahman MA, Mahmud TMM, Abdul Rahman R, Kadir J, Begum MM (2012) Potential Co-application of Burkholderia cepacia, Calcium and Chitosan on enhancement of storage life and quality of papaya fruits. Pertanika J Trop Agric Sci 35(3):439

Ranasinghe P, Ranasinghe P, Abeysekera WKM, Premakumara GS, Perera YS, Gurugama P, Gunatilake SB (2012) In vitro erythrocyte membrane stabilization properties of *Carica papaya* L. leaf extracts. Pharmacogn Res 4(4):196

Rimberia FK, Ombwara FK, Mumo NN, Ateka EM (2018) Genetic Improvement of papaya (*Carica papaya* L.). In: Jameel MA, Jain SM, Johnson DV (eds) Advances in plant breeding strategies: fruits. Springer, Cham, pp 897–928

Sabara P, Vakharia D (2018) Genetic diversity study in papaya (*Carica papaya* L.) cultivars using RAPD and ISSR markers. Indian J Biotechnol 17:101–109

Sabbadini S, Capocasa F, Battino M, Mazzoni L, Mezzetti B (2021) Improved nutritional quality in fruit tree species through traditional and biotechnological approaches. Trends Food Sci Technol 117:125–138

Saeed F, Arshad MU, Pasha I, Naz R, Batool R, Khan AA, Nasir MA, Shafique B (2014) Nutritional and phyto-therapeutic potential of papaya (*Carica papaya* Linn.): an overview. Int J Food Prop 17(7):1637–1653

Salinas G, Luna-Cavazos M, Soto-Estrada A, Garcia-Perez E, Pérez-Vázquez A, Córdova-Téllez L (2022) Distribution and eco-geographic characterization of *Carica papaya* L. native to Mexico. Genet Resour Crop Evol 69(1):99–116

Santa-Catarina R, Vettorazzi JCF, Cortes DFM, Santana JGS, Poltronieri TPDS, Miranda DP, Pirovani AAV, Moreira NF, Vivas M, Pereira MG (2020) Phenotypic characterization of recurrent selection S1 Papaya (*Carica papaya* L.) families by multivariate approach. Euphytica 216(7):1–13

Sarker MMR, Khan F, Mohamed IN (2021) Dengue fever: therapeutic potential of *Carica papaya* L. Leaves. Front Pharmacol 12:610912

Serra S, De Simeis D, Brenna E (2016) Lipase mediated resolution of cis-and trans-linalool oxide (pyranoid). J Mol Catalys B Enzym 133:S420–S425

Sharma A, Bachheti A, Sharma P, Bachheti RK, Husen A (2020) Phytochemistry, pharmacological activities, nanoparticle fabrication, commercial products and waste utilization of Carica papaya L.: A comprehensive review. Curr Res Biotechnol 2:145–160

Sharmin MR, Islam MN, Alim MA (2016) Shelf-life enhancement of papaya with aloe vera gel coating at ambient temperature. J Bang Agric Univ 13(1):131–136

Silva CA, Nascimento AL, Ferreira JP, Schmildt O, Malikouski RG, Alexandre RS, Ferreguetti GAN, Schmildt ER (2017) Genetic diversity among papaya accessions. Afr J Agric Res 12(23): 2041–2048

Singh RN (1964) Papaya breeding – a review. Indian J Hortic 21(2):148–154

Singh V, Rizvi A, Sara US (2021) Standardization and phytochemical screening of *Carica papaya* seeds. Res J Pharm Technol 14(9):4540–4546

Sondur SN, Manshardt RM, Stiles JI (1996) A genetic linkage map of papaya based on randomly amplified polymorphic DNA markers. Theor Appl Genet 93(4):547–553

Suvalaxmi P, Ranjan RG, Kumar DD (2019) Molecular and morphological assessment of papaya (*Carica papaya*). Res J Biotechnol 14:63–70

Swingle CF (1947) The Peruvian cooking papaya, *Carica monoica,* a promising new fruit and vegetable for the United States corn belt. Proc Am Soc Hortic Sci 49:137–138

Teng WC, Chan W, Suwanarusk R, Ong A, Ho HK, Russell B, Rénia L, Koh HL (2019) In vitro antimalarial evaluations and cytotoxicity investigations of *Carica papaya* leaves and carpaine. Nat Prod Commun 14:33–36

Trujillo E, Gámez Vázquez AJ, Avila Perches MA, Palemón Alberto F, Hernández Ruiz J (2018) Potential geographical distribution of papaya wild cultivated in Mexico. Rev Mex Cienc Agrícolas 9(7):1377–1388

Vincent L, Anushma PL, Vasugi C, Rekha A, Shiva B (2019) Genetic resources of tropical fruits. In: Rajasekharan PE, Ramanatha Rao V (eds) Conservation and utilization of horticultural genetic resources. Springer, Singapore, pp 79–116

Wang J, Chen C, Na JK, Yu Q, Hou S, Paull RE, Moore PH, Alam M, Ming R (2008) Genome-wide comparative analyses of microsatellites in papaya. Trop Plant Biol 1(3):278–292

Zhou Z, Ford R, Bar I, Kanchana-Udomkan C (2021) Papaya (*Carica papaya* L.). Flavour profiling. Genes 12(9):1416

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. Plant Genome 1(1):5–20

Zillo RR, da Silva PPM, de Oliveira J, da Glória EM, Spoto MHF (2018) Carboxymethylcellulose coating associated with essential oil can increase papaya shelf life. Sci Hortic 239:70–77

Zoelen DJ, Egmond MR, Pieters RJ, Liskamp RM (2007) Synthesis and evaluation of TAC-based inhibitors of papain as mimics of cystatin B. Chembiochem 8(16):1950–1956

Avocado: Agricultural Importance and Nutraceutical Properties

A. Talavera, J. J. Gonzalez-Fernandez, A. Carrasco-Pancorbo, L. Olmo-García, and J. I. Hormaza

Contents

Abstract

The avocado is a subtropical evergreen tree crop originated in Mesoamerica presently cultivated worldwide in more than 60 countries. Avocado fruits were consumed by Native American cultures as early as 10,000 years ago, and they are experiencing increasing popularity globally as highly nutritious and healthy food. In this chapter, we review the agricultural importance of the crop, compositional

A. Talavera · J. J. Gonzalez-Fernandez · J. I. Hormaza (✉)
Instituto de Hortofruticultura Subtropical y Mediterranea La Mayora (IHSM La Mayora-UMA-CSIC), Malaga, Spain
e-mail: jorgegonzalez-fernandez@eelm.csic.es; ihormaza@eelm.csic.es

A. Carrasco-Pancorbo · L. Olmo-García
Department of Analytical Chemistry, Faculty of Sciences, University of Granada, Granada, Spain
e-mail: alegriac@ugr.es; luciaolmo@ugr.es

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_40

profile, genetic diversity, and molecular tools developed in avocado breeding. The avocado fruit and some of its byproducts (peels, seeds) show nutraceutical properties that can be different depending on the variety and the preharvest and postharvest management approaches.

Keywords

Avocado · Genomics · Nutraceutics · Lauraceae · *Persea americana*

1 Introduction

1.1 Agricultural Importance of Avocado

The avocado (*Persea americana* Mill.) is a subtropical evergreen woody perennial tree crop, member of the family Lauraceae in the order Laurales. The center of origin of avocado can be located in Central America, in a wide geographical region that expands from the highlands of Mexico, Guatemala, and Honduras to the Pacific coast (Popenoe et al. 1997). According to archaeological findings, native cultures in Mexico already consumed avocados at least 10,000 years ago, as remains dating back to that time have been found in the Coxcatlan Cave, in Tehuacan (state of Puebla) (Knight 2002). Originally, the word avocado derives from the ancient nahuatl word "ahuacatl" that the Spaniards transformed into "aguacate" from which the English word "avocado" transliterated. As other trees native of the Americas that produce large fruits, it is likely that the avocado coevolved with the presently extinct American megafauna that probably was the main disperser of the fruits before humans arrived to the continent (van Zonneveld et al. 2018). From its center of origin in Central America, avocado was probably dispersed to South America in pre-Columbian times. In fact, the first description of avocado after the European arrival to America was made in Colombia in 1519 (Fernandez de Enciso 1519). Soon, the European explorers dispersed avocados to other regions of the world, and the first description outside the Americas was made in 1576 by the botanist Charles de l'Écluse (Carolus Clusius) (Clusius 1576), who observed a blooming avocado tree in a botanical garden in Valencia (Spain), established by Joan Plaza, a professor of "herbs and other simple medicines" at the University of Valencia (Lopez-Terrada 2011). Clusius named the tree Persea due to its resemblance to a traditional North African tree with that name (*Mimusops laurifolia*, Sapotaceae) (Schroeder 1977); this will later become the name of the genus (Miller 1754).

At least eight botanical varieties or subspecies that have evolved in different edaphoclimatic conditions and geographically isolated from each other are usually recognized in the species *Persea americana*. Three of them, also called horticultural races, have agronomic importance (Schaffer et al. 2013): West Indian (*P. americana* var. *americana*), from the lowland tropics; Guatemalan (*P. americana* var. *guatemalensis*), from the valleys of the Central American mountain ranges; and Mexican (*P. americana* var. *drymifolia*), from the tropical highlands of Southern

Mexico. Thus, the three horticultural races differ mainly in botanical characteristics and edaphoclimatic preferences. The Guatemalan and Mexican subspecies are better adapted to cooler subtropical climates although the Guatemalan subspecies is more vulnerable to low temperatures than the Mexican subspecies. The West Indian subspecies is more adapted to tropical climates. The three horticultural races are sexually intercompatible, and, in fact, most of the presently cultivated avocado commercial varieties in subtropical and Mediterranean climates are interracial Mexican x Guatemalan hybrids. Avocado is presently cultivated in most countries with subtropical and tropical climates all over the world and is grown commercially in more than 60 countries worldwide.

Avocado international market and trading have seen an exponential increase in the last couple of decades, and global avocado world production in 2020 reached over eight million tons (FAO 2022). Mexico, Dominican Republic, Peru, Indonesia, Colombia, and Brazil contribute most of the production. Mexico is the most important avocado-producing country with approximately 30% of the global production (over two million tons in 2022). Most avocados in the international trade are exported from those main producing countries to big consumer regions, mainly USA and Western Europe, while the rest is predominantly consumed in local markets. Thus, the avocado can be considered as a staple fruit in most countries with tropical and subtropical climates, especially in the Americas, but also in some African and Asian countries. Local landraces that usually do not reach the export markets are very popular in most of those countries. Encouraging smallholder farmers to grow avocados together with other crops could, thus, help to reduce malnutrition and help to increase food security (Hakizimana and May 2018).

1.2 Relationship with Other Species and Wild Relatives

The Laurales, in addition to the Canellales, Magnoliales, and Piperales, form the magnoliid clade that is considered as a sister clade to the eudicot and monocot angiosperms (APG IV 2016; Chase et al. 2016). The order Laurales includes seven families: Atherospermataceae, Calycanthaceae, Gomortegaceae, Hernandiaceae, Lauraceae, Monimiaceae, and Siparunaceae. The Lauraceae is a monophylic family with more than 50 genera and between 2500 and 3000 species with a worldwide distribution, mainly in tropical and subtropical climates. Most of the species of the family are evergreen woody trees or shrubs, and, in addition to avocado, a few other species in the Lauraceae have agronomic and/or economic interest. Among them, the most important are spices, such as the camphor (*Cinnamomum camphora* [L.] J. Presl), the cinnamon (*Cinnamomum verum* J. Presl), or the laurel (*Laurus nobilis* L.); timber trees such as those of the genera *Chlorocardium, Eusideroxylon*, *Mezilaurus, Nectandra*, *Ocotea*, and *Phoebe*; and ornamental trees such as *Persea indica*. About 150 species are recognized in the genus *Persea*, 70 of which are distributed in America and 80 in Southeastern Asia, and a single species, *P. indica*, is endemic to the Macaronesian islands in the Atlantic Ocean. The genus possibly first evolved in African Gondwanaland, from where it dispersed to North America and Asia through

Europe; by the Paleogene, the genus reached South America via Antarctica, and it was reunited thanks to the land bridge that during the late Neocene connected North and South America.

Three subgenera are usually considered in the genus *Persea*: *Machilus*, *Persea*, and *Eriodaphne*, although *Machilus*, restricted to Asia, is treated as a separate genus by many authors. The 70 American species of *Persea* are divided in two subgenera (Kopp 1966): *Eriodaphne* (primarily with a South American origin) and *Persea* (primarily with a Central American origin). The subgenus *Persea* includes at least three species, although additional species are added by some authors (Chanderbali et al. 2013): *P. pallescens*, *P. schiedeana*, and *P. americana*, the avocado.

2 Description on Nutritional Constituents

2.1 Products/Parts with Nutritional Interest

Avocado fruits develop from flowers that are produced as panicles of cymes. A mature avocado tree may produce more than a million flowers, although less than 1% remain on the tree at harvest, which results in a very low fruit-to-flower ratio (Alcaraz and Hormaza 2021). The fruit of avocado is botanically a pyriform or globose fleshy drupe that contains one seed surrounded by the pericarp. The pericarp is formed by the outer skin (exocarp) with a variable thickness and color depending on the horticultural race and the variety, the pulp (mesocarp) that constitutes the edible part, and a thin endocarp, next to the two papery seed coats (Cummings and Schroeder 1942). The pericarp color and texture are highly variable, and the color at ripening can be black, brown green, or purple with intermediate intensities of each one depending on the genotype. The fruit shape and size are also variable and can range from spherical to different pear or egg shapes.

The avocado fruit developmental process can be divided in two main phases: fruit maturation that takes place on the tree and that, depending on both the environmental conditions and the genotype, can last between 6 and 18 months and postharvest ripening that takes place after the detachment of the fruit from the tree and during which the mesocarp softens changing the organoleptic characteristics of the fruit. The avocado fruit is climacteric, and, consequently, during the ripening process after harvest, respiration levels related to ethylene production increase. The fruits are harvested when physiological maturity (the ability to ripen after harvest) is reached. Edible ripeness takes place several days after harvesting; the length of the phase between physiological and edible maturity is variable depending on the variety, the time of harvesting, and the conservation temperature after harvesting. The visual determination of the stage of physiological maturity is difficult in avocado, since no clear external changes are observed during this process. This topic has been addressed by a number of works (Stahl 1933; Lee et al. 1983; Landahl et al. 2009) resulting in the development of different approaches to accurately assess avocado physiological fruit maturity; among the most used is the correlation between dry matter and oil contents (Gamble et al. 2010).

As avocado fruits develop on the tree, total mesocarp oil content increases, and, thus, oil content in the fruit can be used to establish the different fruit developmental stages in order to decide when physiological maturity has been reached (Lee 1981). Avocado is one of the few examples in which high amounts of triacylglycerols are accumulated in the mesocarp of the fruit (Kilaru et al. 2015). Although in the 1920s in California a minimum empirical value of 8% oil content was established for harvesting, this value was later considered as too low. In any case, determination of oil content is a cumbersome process, and, consequently, alternative determination methods were necessary. During avocado fruit development, as oil content increases, since oil replaces the water present in the fruit, water content decreases concomitantly. Then, during this process, the increase in both oil and dry matter contents is correlated. Consequently, measurement of dry matter (defined as the solid content of the fruit minus its water content) is an indirect indicator of oil content and, thus, of maturity and a reliable indicator of flavor. Dry matter measurements are easy to perform either using microwave or other ovens to dry avocado fruit slices or, more recently, using near-infrared (NIR) technologies. The minimum dry matter value for harvesting is variable depending on the variety and the region of production, but the standard value for 'Hass' avocado ranges from 21% to 23%.

Most avocados are consumed fresh but also in salsas (such as guacamole), soups, and other elaborations, such as in Brazil and different Asian countries in which avocados are consumed sweet mixed with sugar and condensed milk, in milk shakes or as ice cream. In addition, increasing interest is being put into using the edible oil extracted from the avocado flesh in addition to the more traditional use of the avocado oil in the cosmetic and beauty industries.

Although the main interest of this crop is the fruit consumed mainly fresh, in some cases, the leaves of the avocado trees (mainly of genotypes of the Mexican race which have a distinctive anise aroma) are also used in some traditional cuisines. Also, avocado leaves are used in some regions in traditional medicine against coughs or skin bruising. Other parts of the avocado fruits after processing (such as for guacamole production) that are usually considered as waste (peels, seeds, paste) could also be a source of interesting bioactive substances (Dalle Mulle Santos et al. 2016; Araújo et al. 2018). Exploiting the phytochemical properties of those byproducts will increase the added value of the avocado industry helping to improve sustainability of avocado production worldwide.

2.2 Detailed Chemical Composition

Avocado is a calorie-rich fruit with high levels of unsaturated fatty acids. The avocado fruit is particularly rich in ascorbic acid (vitamin C), vitamin B6, β-carotene, vitamin E, and potassium, and it also contains phytosterols and carotenoids, such as lutein (which represents about 70% of the carotenoids of avocado [Dreher and Davenport 2013]) and zeaxanthin. For a standard 'Hass' avocado, we can consider that about 72% of the flesh is water, although this percentage is variable depending on the location and on the stage of fruit development. Avocado

composition of the edible flesh for the most important macro- and micronutrients per 100 g of 'Hass' avocados is shown in the following Table 1 (USDA 2018).

Compared to other vegetables with high oil content, the avocado fruit shows high levels of monounsaturated fatty acids (such as oleic acid), whereas the levels of polyunsaturated fatty acids (such as linoleic and linolenic acids) are low. It also contains small proportions of additional fatty acids (stearic, palmitic, myristic, or arachidonic acids).

For 100 g, it can be considered that 2.13 g corresponds to saturated fatty acids, whereas 9.8 g corresponds to monounsaturated fatty acids (about 70% of total fats) and 1.82 g to polyunsaturated fatty acids (USDA 2018).

We can highlight some of the main health-related nutrients of avocado fruits as follows:

- Relevant levels of monounsaturated fatty acids.
- Relevant levels of dietary fiber, of which about 70% correspond to insoluble fibers and 30% to soluble fibers (Marlett and Cheung 1997).
- Low sugar content compared to other fruits, mainly in the form of the heptose *D*-mannoheptulose and its reduced form, the sugar alcohol perseitol.

Table 1 Most important macro- and micronutrients per 100 g of 'Hass' avocado

Water (g)	72.3
Macronutrients (g)	
Lipids	15.4
Proteins	1.96
Total carbohydrates	8.64
Dietary fiber	6.80
Minerals (mg)	
Calcium	13.0
Copper	0.17
Iron	0.61
Magnesium	29
Phosphorous	54
Manganese	0.15
Potassium	507
Selenium	0.40
Sodium	8
Zinc	0.68
Vitamins	
Niacin (mg)	1.91
Pantothenic acid (mg)	1.46
Riboflavin (mg)	0.14
Thiamin (mg)	0.075
Vitamin A (IU)	147
Vitamin B6 (mg)	0.29
Vitamin C, ascorbic acid (mg)	8.80

- High potassium and low sodium content, which can help to control blood pressure levels.
- High magnesium levels.
- High levels of antioxidant C and E vitamins.
- High levels of B vitamins.
- High levels of phytosterols, mainly β-sitosterol, followed by campesterol and stigmasterol (Duester 2001).

Significant changes can be observed in the concentrations of some of these compounds along the harvesting season and also during ripening after harvest (Lu et al. 2009; Hurtado-Fernández et al. 2016; Serrano-García et al. 2022). Hence, the levels of saturated fat decrease, whereas those of monounsaturated oleic acid increase (Slater et al. 1975; Lu et al. 2009). The levels of carotenoids also increase along the avocado harvesting season (Lu et al. 2009).

Avocado fruits also show high levels of certain types of nonnutritive compounds, such as alkanols, amino acids, carotenoids, or phenolics, which account for some of their organoleptic properties and that could also be relevant in improving human health.

The first works in which powerful tools and ambitious analytical methods were used to characterize the avocado pulp exhaustively – considering many substances not described in the food composition databases – were performed by Hurtado-Fernández et al. (2011a, b). After that, other works have been published addressing the study of the substances present not only in the avocado pulp but also in the avocado peel and seeds (López-Cobo et al. 2016; Figueroa et al. 2018a, b).

2.3 Medicinal/Physiological Properties and Functions in Relation to Human Health

Avocado is considered as an excellent source of different macro- and micronutrients, and significant health benefits of avocado pulp and oil consumption have been described in different works (Bhuyan et al. 2019). Among them, we can include lowering risks of cardiovascular diseases, cataracts and other age-related macular degenerations, metabolic syndromes, prostatic hypertrophy, and prostate and other tumors, as well as blood cholesterol regulation, weight management, diabetes control, anti-inflammatory effects, and prevention and treatment of osteoarthritis, among others (Salazar et al. 2005; Dreher and Davenport 2013; Bhuyan et al. 2019). Those benefits are probably related to the presence of different phytochemicals in the avocado fruits including phenolic compounds, chlorophylls, carotenoids, or anthocyanins (Ashton et al. 2006; Ding et al. 2007) as well as to the high levels of monounsaturated and low levels of saturated fats that result in a similar healthy oil profile to that of the olive oil (Bhuyan et al. 2019).

Most of the studies performed in the avocado fruit have been focused on the pulp, consumed either fresh or processed as oil, which often is cold-pressed, and, consequently, most of the bioactive compounds are preserved (dos Santos et al. 2014). In

fact, there is a huge potential to improve the use of avocado as a high-quality oil, since generally all the breeding and selection of varieties in this crop have been focused on fruit quality for the fresh market and the diversity present in this species for oil production has not been studied in detail. Moreover, displacing the production of cooking oil from annual plants to perennial crops could have an overall impact on sustainability, similar to the case of the olive tree in the Mediterranean region, helping to reduce soil erosion and fertilizer and pesticide runoff while providing a more sustainable agricultural landscape (Bost et al. 2013). In some markets, fruits of West Indian and of hybrid Guatemalan x West Indian varieties that show a lower oil and calorie content than 'Hass' are marketed as "light avocados."

In addition to the monounsaturated acids, recent interest has been devoted to other fatty acid derivatives present in the avocado fruit, such as lauraceous acetogenins that show different interesting bioactive properties (Rodríguez-López et al. 2017; Colin-Oviedo et al. 2022). These substances were first reported in avocado leaves several decades ago (Chang et al. 1975).

A recent work (Sánchez-Quezada et al. 2021) shows that nutraceutical properties of avocado seeds (which are responsible for about 16–22% of the avocado fruit weight and are often a byproduct from the guacamole industry) correlate positively with the fruit ripening process; as the fruit ripens, seed moisture decreases, and the antioxidant capacity increases due to an increasing concentration of phenolics.

2.4 Cultural Methods for Nutraceutical Improvement

Although avocado international trade is mainly based on a single variety, 'Hass' (Fig. 1), which allows the availability in the markets of an externally very

Fig. 1 Fruits of the avocado variety 'Hass' hanging on a tree before harvesting

homogeneous fruit all year round from different origins, pre- and postharvest management factors play a relevant role on 'Hass' avocado fruit quality and nutritional composition. Preharvest factors include fruit origin, which, in turn, is mostly related to edaphoclimatic conditions, especially soil type and chemical status, light intensity, relative humidity, temperature fluctuations during fruit development, harvest date along the harvesting season, and agronomic management, especially irrigation regimes and fertilization programs. Postharvest management includes handling at harvesting time, during transport to cooling facilities, and at packing operations, storage length and conditions, and processing and transport to final destination markets.

Although very few works have addressed the influence of all those factors on avocado nutraceutical properties, the understanding and control of those factors are highly relevant in order to optimize avocado fruit quality and composition. One example is the length of the harvesting season of fruits derived from the same flowering period that, in some avocado-producing regions, can be extended up to 6 months with differences in fruit composition depending on the harvest time (Hurtado-Fernández et al. 2016) or the geographical origin (Pedreschi et al. 2022). In this sense, differences in avocado fruit composition in the same variety related to the growing area, ripening stage, and postharvest fruit storage conditions have also been described (Donetti and Terry 2014).

Regarding crop management, key aspects that can explain heterogeneous ripening in avocado include mainly irrigation and pruning of fertilization practices (Rivera et al. 2017). In this sense, a recent work combining proteomic and metabolomic studies revealed clear differences after heat treatments between avocados from early and mid-harvesting seasons; those differences seem to depend on the fruit physiological stage at harvest (Gavicho-Uarrota et al. 2019). Heat treatments can improve the homogeneity of the ripening process by increasing the amount of soluble sugars (such as galactose or sucrose) and of some stress-related enzymes. Regarding postharvest management, a careful fruit handling has to be performed avoiding not only mechanical injuries but also controlling environmental conditions that could alter normal fruit maturation and avoiding fruit disorders or postharvest diseases that could affect final fruit quality and nutraceutical properties.

Although, as discussed above, differences can be found in fruit composition in the same cultivar depending on the edaphoclimatic conditions and pre- and postharvest management procedures, the overwhelming use of 'Hass' in the international avocado market is a limiting factor to exploit the natural diversity present in this crop. Some studies have shown differences among avocado varieties in fruit composition (Hurtado-Fernandez et al. 2015; Di Stefano et al. 2017); this will allow in the future to combine the production of different varieties increasing the avocado production period; in some cases, such as in Spain and other countries with a Mediterranean climate, it is possible to produce avocados all year round with a combination of four or five varieties. In other countries with subtropical climates (Colombia or Mexico), different flowering periods can take place all year round allowing the production of fruits from the same variety during the whole year.

On the other hand, in countries with Mediterranean climates where the pressure of pests and diseases is low compared to regions with tropical high humidity conditions, avocado organic production is relatively easy to perform with the consequently improvement of healthy food production due to the lack of application of toxic chemicals for pest and disease control.

3 Molecular Tools for Genetic Improvement of Nutraceuticals

3.1 Limitations of Conventional Breeding and Rational for Next-Generation Molecular Breeding

The avocado is highly heterozygous and shows a long juvenile period and, as stated before, with an extremely low fruit set due to massive flower abscission and immature fruit drop (Alcaraz and Hormaza 2021). Avocado is characterized by the presence of a synchronous protogynous dichogamy system in which each hermaphrodite avocado flower opens twice during two consecutive days, first functionally as a female flower and, during the second opening, as a functionally male flower. Avocado genotypes have been traditionally classified in two different types (A or B) depending on whether in the first day of the flowering cycle the flowers open in the female stage in the morning (type A cultivars) or in the afternoon (type B cultivars). This flowering behavior promotes outcrossing, since self-pollination in the same cultivar can only occur during the limited number of hours in which closing female flowers in the first day of the flowering cycle and opening male flowers in the second day of the flowering cycle can be found on the same inflorescence, tree, or cultivar. The length of this overlap is highly dependent on temperature and on the avocado genotype.

Conventional breeding programs have been moderately successful to release new avocado varieties and rootstocks, although the majority of the most important commercial cultivars worldwide are derived from open pollinations and spontaneous interracial hybridizations. In general, the objectives of avocado breeding strategies have focused on (i) high consistent yield, (ii) high fruit quality, (iii) expanded harvesting season, and (iv) longer shelf life (Lahav and Lavi 2013). The most important avocado cultivar present in the international markets is ‘Hass’ that produces fruits with a black rough epicarp, followed, to a much lesser extent, by varieties such as ‘Fuerte’ that produce fruits with a green smooth epicarp. This has not always been the case, since ‘Fuerte’ was the leading avocado variety in the international markets until the 1970s. Several cultivars similar to ‘Hass’ in appearance have been released during the last decades, and some of them are cultivated in some regions, such as ‘Lamb Hass’, ‘Carmen’ (a mutation of ‘Hass’), or ‘Maluma’. Other interesting cultivars that ripen in green are ‘Reed’ (type A, as ‘Hass’) and ‘Bacon’, ‘Edranol’, or ‘Zutano’ which are type B and are used as pollinizers for ‘Hass’.

In spite of the growing importance of avocado production worldwide, significant bottlenecks that limit breeding and development of new high-quality avocado varieties still remain worldwide. Among the main limiting factors, we can include the lack or scarce availability of molecular resources, the absence of adequate phenotypic data for many cultivars and environments, and the very limited germplasm used in most avocado breeding programs. There is an increasing demand to develop new high-quality avocado cultivars, since about 90% of the world avocado production is based on 'Hass', a variety that was developed in the 1920s in California (USA) from a chance seedling and, in fact, all presently cultivated 'Hass' trees derive clonally from that single tree patented in 1935 (Crane et al. 2013). Therefore, any pest or disease that affects this cultivar could threaten avocado production worldwide. Moreover, the present scenario of climate change requires adapting the breeding programs to rapidly changing environmental conditions. As in other crops, it is expected that the increasing availability of molecular data derived from high-throughput sequencing technologies will speed up avocado breeding in the next years.

3.2 Molecular Genetics and Genomics of Nutraceuticals

The avocado is a diploid ($2n = 2x = 24$) species with a haploid genome size of 896 Mb (Arumuganathan and Earle 1991). Different molecular markers have been used in this crop for fingerprinting, paternity assessment, diversity and phylogenetic analyses, development of genetic linkage maps or screening, and selection of traits of interest in avocado breeding programs. As with other crops, in avocado isozymes were the first genetic markers used (Torres and Bergh 1980) followed more recently by the development of different DNA-based markers.

Early markers based on DNA included minisatellites (Lavi et al. 1991), variable number of tandem repeats (VNTRs) (Mhameed et al. 1996), random amplified polymorphic DNA (RAPDs) (Fiedler et al. 1998), and restriction fragment length polymorphism (RFLP) (Furnier et al. 1990; Davis et al. 1998). Later, codominant and highly polymorphic single sequence repeats (SSRs) or microsatellites were also specifically developed in avocado and widely applied for cultivar fingerprinting and genetic diversity studies (Sharon et al. 1997; Schnell et al. 2003; Ashworth and Clegg 2003; Ashworth et al. 2004; Borrone et al. 2007; Alcaraz and Hormaza 2007; Gross-German and Viruel 2013; Guzmán et al. 2017; Boza et al. 2018; Ge et al. 2019b; Juma et al. 2021; Ruiz-Chután et al. 2022). The more recent development of high-throughput new-generation sequencing technologies has enabled the use of single-nucleotide polymorphism (SNP) markers that are increasingly being considered as the preferable type of marker for genetic studies in different crops for diverse objectives such as the construction of genetic linkage maps, the characterization of quantitative trait loci (QTL), association studies with traits of agronomic interest, marker-assisted selection (MAS), or genomic selection (GS) (Scheben et al. 2017; Le Nguyen et al. 2019). SNPs show several advantages that are especially relevant in woody perennial crops, such as avocado, in which breeding programs are

characterized by long generation times: the possibility of developing a large number of markers at increasingly reduced costs, their bi-allelism that results in a high accuracy in the analyses, or their high replicability.

New-generation sequencing approaches in avocado have so far mostly been focused on transcriptome analyses and SNP development. In general, these studies have been applied to study disease resistance, plant development, stress tolerance, and genetic diversity (Ibarra-Laclette et al. 2015; Kilaru et al. 2015; Liu et al. 2018; Vergara-Pulgar et al. 2019; Ge et al. 2019a; Kuhn et al. 2019a, b; Ge et al. 2019c; Rubinstein et al. 2019; Talavera et al. 2019; Chabikwa et al. 2020; Pérez-Torres et al. 2021; Kämper et al. 2021; Fick et al. 2022; Hernández et al. 2022). However, recently, several avocado nuclear genome sequences have been published. Rendón-Anaya et al. (2019) developed genomic sequence drafts of 'Hass' as well as of a wild accession of the Mexican horticultural race. The 'Hass' assembly was anchored to a genetic map, and the generated genome consisting of 915 scaffolds covered approximately half (46.2%) of the estimated avocado genome. Furthermore, ten genomes representing the three avocado subspecies, as well as Hass-related cultivars and *P. schiedeana*, were resequenced helping to better understand the genetic diversity and the evolutionary history of avocado. Subsequently, another 'Hass' genome reference has been generated by Sharma et al. (2021). In this last work, the reference genome resulted in an assembly of 788 Mb representing approximately 88% of the avocado genome size. However, this assembly was not anchored into scaffolds. These new genomic resources in avocado will probably make a qualitative change in future avocado breeding. But additional efforts are needed to find solutions to new challenges, since the molecular knowledge of avocado lags far behind other crops.

3.3 Molecular Mapping, QTLs, and Gene Identification

The increased availability of saturated genetic maps in many crops has accelerated different breeding programs in recent years. These maps have been extensively applied in QTL mapping, MAS, and comparative and genome structure analyses allowing the identification of a large number of genes through QTL fine mapping studies (Jaganathan et al. 2020). However, in avocado, the number of linkage maps constructed to date is still low. The first avocado genetic linkage map was developed with 50 SSRs, 23 minisatellite DNA fingerprints (DFPs), and 17 RAPD markers, using the progeny from a cross between 'Pinkerton' and 'Ettinger'. This map represented the 12 avocado linkage groups and covered 352.6 cM (Sharon et al. 1997). Later, a new linkage map was constructed by Borrone et al. (2009) using 715 F1 individuals (456 from the cross 'Tonnage' x 'Simmonds' and 259 from the reciprocal cross 'Simmonds' x 'Tonnage'). This map was developed using SSR makers, and 12 linkage groups were also detected, but, in this case, the total map length was longer (1087.4 cM). More recently, a linkage map was constructed using SSRs and single-nucleotide polymorphisms for a 'Gwen' x 'Fuerte' progeny. The

resulting linkage map covered 1044.7 cM distributed along 12 linkage groups. In this last study, a high number of molecular markers located in linkage group 10 showed remarkable association with flowering type, and a marker on linkage group 1 was associated with a QTL related to β-sitosterol fruit levels and a region on linkage group 3 with vitamin E (α-tocopherol) fruit levels. Unfortunately, the limited population size used restricted the possibility of developing a robust linkage map and detecting additional QTLs (Ashworth et al. 2019).

Although the genetic maps generated so far are useful resources, new efforts are needed to develop more saturated genetic maps in avocado in order to optimize and boost breeding programs for specific traits. Several mapping populations have been developed in spite of the difficulty of this crop with a long juvenile period, the low yield obtained after natural or hand pollination, and the dichogamy system. However, larger populations, higher molecular resources, and phenotypic data will be required in order to detect QTLs and develop marker-assisted programs.

As a result of those limitations, the examples of the identification of health-related genes are still scarce in avocado. Those examples include the studies on oil accumulation in the mesocarp of avocado fruits. Kilaru et al. (2015) performed a comparative transcriptome analysis of some lipid metabolic pathways in 'Hass' and compared this with other species showing a similar oil biosynthesis patterns (avocado, oil palm, rapeseed, and castor bean) suggesting the conservation of these metabolic pathways during angiosperm evolution. However, some unique features were also detected in avocado. Similarly, comparative transcriptomic analyses of the mesocarp and seeds revealed higher expression levels of 17 carotenoid biosynthesis genes in the mesocarp than in the seed along five avocado fruit developmental stages (Ge et al. 2019a).

3.4 Genetic Engineering

As in other crops, genetic engineering could play an important role in future avocado breeding programs. Significant advances have been obtained using *Agrobacterium rhizogenes* and *Agrobacterium tumefaciens* (Pliego-Alfaro et al. 2020). Genetic transformation has been successful in a few cases to study some genes that affect different horticultural traits. Examples include the manipulation of the *S*-adenosyl-l-methionine (SAM) hydrolase (SAMase) gene to block ethylene production (Efendi 2003) or the study of genes involved in pathogenesis such as β-1,6-glucanase, chitinase, and the antifungal protein (AFP) gene (Raharjo et al. 2008) or targeting chloroplast RNA to avoid the replication of the avocado sunblotch viroid (ASBVd) in chloroplasts (Perea Arango et al. 2010). Transient gene expression has also been accomplished using biolistics in avocado embryogenic cultures (Chaparro-Pulido et al. 2014). In addition, as in other crops, CRISPR/Cas can make a qualitative change in avocado gene edition in the future. However, efficient regeneration protocols from single cells in this crop are still lacking in order for the approach to be effective.

4 Future Prospects

This chapter has summarized relevant information about avocado, covering aspects such as the agricultural importance of the crop, description of compositional profile, genetic diversity, and molecular tools. Although the composition of this fruit has been described in considerable depth, there is still a long way to go, making use of recent advances in metabolomics, proteomics, sequencing, transcriptomics, and genomics. Thus, in spite of the importance of several compounds present in avocado fruits with nutritional and health benefits, such as carotenoids, lipids, sugars, proteins, minerals, or vitamins, avocado lags behind other crops in the breeding or varieties to take advantage of those excellent properties. Several studies have shown the exceptional nutritional and phytochemical composition of avocado fruits and their potential in the control and prevention of different human diseases, but additional in vitro, in vivo, and clinical studies are needed to understand the mechanisms of action of avocado phytochemicals in order to use this fruit in therapeutic and nutritional applications. However, recent advances in sequencing and transcriptomics together with gene editing can make a qualitative change to take advantage of the still underused genetic diversity present in this crop. It is expected that additional genes coding for important nutraceutical properties of avocado will be identified in the next years. This knowledge is crucial to use molecular breeding to improve classical breeding approaches as well as to provide appropriate material for the generation of DNA-edited plants with desired traits of interest.

5 Conclusion

The fruits of the avocado including several of their byproducts show interesting nutraceutical properties that can be different depending on the genotype and the preharvest and postharvest management of the crop.

References

Alcaraz ML, Hormaza JI (2007) Molecular characterization and genetic diversity in an avocado collection of cultivars and local Spanish genotypes using SSRs. Hereditas 144(6):244–253

Alcaraz ML, Hormaza JI (2021) Fruit set in avocado: pollen limitation, pollen load size, and selective fruit abortion. Agronomy 11(8):1603

APG IV (2016) An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc 181:1–20

Araújo RG, Rodriguez-Jasso RM, Ruiz HA, Pintado MME, Aguilar CN (2018) Avocado by-products: nutritional and functional properties. Trends Food Sci Technol. 80:51–60

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

Ashton OBO, Wong M, McGhie TK, Vather R, Wang Y, Requejo-Jackman C, Ramankutty P, Woolf AB (2006) Pigments in avocado tissue and oil. J Agric Food Chem 54:10151–10158

Ashworth VETM, Clegg MT (2003) Microsatellite markers in avocado (*Persea americana* mill.): genealogical relationships among cultivated avocado genotypes. J Hered 94(5):407–415

Ashworth VETM, Kobayashi MC, De La Cruz M, Clegg MT (2004) Microsatellite markers in avocado (*Persea americana* mill.): development of dinucleotide and trinucleotide markers. Sci Hortic 101(3):255–267

Ashworth VE, Chen H, Calderón-Vázquez CL, Arpaia ML, Kuhn DN, Durbin ML, Tommasini L, Deyett E, Jia Z, Clegg MT, Rolshausen PE (2019) Quantitative trait locus analysis in avocado: the challenge of a slow-maturing horticultural tree crop. J Am Soc Hortic Sci 144(5):352–362

Bhuyan DJ, Alsherbiny MA, Perera S, Low M, Basu A, Devi OA, Barooah MS, Li CG, Papoutsis K (2019) The odyssey of bioactive compounds in avocado (*Persea americana*) and their health benefits. Antioxidants 8(10):426

Bost JB, Smith NJH, Crane JH (2013) History, distribution and uses. In: Schaffer B, Wolstenholme BN, Whiley AW (eds) The avocado: botany, production, and uses. CAB International 2013

Borrone WJ, Schnell RJ, Viola HA, Ploetz RC (2007) Seventy microsatellite markers from *Persea americana* Miller (avocado) expressed sequences tags. Mol Ecol Notes 7:439–444

Borrone JW, Brown JS, Tondo CL, Mauro-Herrera M, Kuhn DN, Violi HA, Sautter RT, Schnell RJ (2009) An EST-SSR-based linkage map for *Persea americana* mill. (avocado). Tree Genet Genomes 5(4):553–560

Boza EJ, Tondo CL, Ledesma N, Campbell RJ, Bost J, Schnell RJ, Gutiérrez OA (2018) Genetic differentiation, races and interracial admixture in avocado (*Persea americana* mill.), and *Persea* spp. evaluated using SSR markers. Genet Resour Crop Evol 65(4):1195–1215

Chabikwa TG, Barbier FF, Tanurdzic M, Beveridge CA (2020) *De novo* transcriptome assembly and annotation for gene discovery in avocado, macadamia and mango. Sci Data 7:9

Chanderbali AS, Soltis DE, Soltis PS, Wolstenholme BN (2013) Taxonomy and botany. In: Schaffer B, Wolstenholme BN, Whiley AW (eds) The avocado: botany, production, and uses. CAB International 2013

Chang C-F, Isogai A, Kamikado T, Murakoshi S, Sakurai A, Tamura S (1975) Isolation and structure elucidation of growth inhibitors for silkworm larvae from avocado leaves. Agric Biol Chem 39:1167–1168

Chaparro-Pulido CA, Montiel MM, Palomo-Ríos E, Mercado JA, Pliego-Alfaro F (2014) Development of an efficient transient transformation protocol for avocado (*Persea americana* mill.) embryogenic callus. In Vitro Cell Dev Biol 50:292–298

Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, Soltis DE, Mabberley DJ, Sennikov AN, Soltis PS, Stevens PF (2016) An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc 181(1):1–20

Clusius C (1576) Rariorum aliquot Stirpium per Hispanias Observatarum Historia. Ex officina Christophori Plantini, Architypographi Regij, Antverpiae

Colin-Oviedo A, Garza-Aguilar SM, Marín-Obispo LM, Rodríguez-Sánchez DG, Trevino V, Hernández-Brenes C, Díaz de la Garza RI (2022) The enigmatic aliphatic acetogenins and their correlations with lipids during seed germination and leaf development of avocado (*Persea americana* mill.). Front Plant Sci 3(13):839326

Crane JH, Douhan G, Faber BA, Arpaia ML, Bender GS, Balerdi CF, Barrientos-Priego AF (2013) In: Schaffer B, Wolstenholme BN, Whiley AW (eds) Cultivars and rootstocks in the avocado: botany, production, and uses. CABI, Wallingford, UK, pp 1–9

Cummings K, Schroeder CA (1942) Anatomy of the avocado fruit. California Avocado Society Yearbook 27:56–64

Dalle Mulle Santos C, Pagno CH, Costa TMH, Faccin DJL, Flores SH, Cardozo NSM (2016) Biobased polymer films from avocado oil extraction residue: production and characterization. J Appl Polym Sci 133(37):1–9

Davis J, Henderson D, Kobayashi M, Clegg MT, Clegg MT (1998) Genealogical relationships among cultivated avocado as revealed through RFLP analysis. J Hered 89:319–323

Di Stefano V, Avellone G, Bongiorno D, Indelicato S, Massenti R, Lo Bianco R (2017) Quantitative evaluation of the phenolic profile in fruits of six avocado (*Persea americana*) cultivars by ultra-high-performance liquid chromatography-heated electrospray-mass spectrometry. Int J Food Prop 20(6):1302–1312

Ding H, Chin YW, Kinghorn AD, D'Ambrosio SM (2007) Chemopreventive characteristics of avocado. Fruit. Semin Cancer Biol 17:386–394

Donetti M, Terry LA (2014) Biochemical markers defining growing area and ripening stage of imported avocado fruit cv. Hass. J Food Compos Anal 34(1):90–98

Dreher ML, Davenport AJ (2013) Hass avocado composition and potential health effects. Crit Rev Food Sci Nutr 53(7):738–750

Duester KC (2001) Avocado fruit is a rich source of beta-sitosterol. J Am Diet Assoc 101(4):404–405

Efendi D (2003) Transformation and cryopreservation of embryogenic avocado (*Persea americana* mill.) cultures. PhD dissertation,. University of Florida, Gainesville

FAO (2022) Statistics division of food and agriculture organization of the United Nations (FAOSTAT). https://www.fao.org/faostat/en/#data/QCL. Accessed 1 July 2022

Fernandez de Enciso M (1519) Suma de Geografia. Sevilla, Spain

Fick A, Swar V, Backer R, Bombarely A, Engelbrecht J, van den Berg N (2022) Partially resistant avocado rootstock Dusa® shows prolonged upregulation of nucleotide binding-leucine rich repeat genes in response to *Phytophthora cinnamomi* infection. Front Plant Sci 13:793644

Fiedler J, Bufler G, Bangerth F (1998) Genetic relationships of avocado (*Persea americana* mill.) using RAPD markers. Euphytica 101:249–255

Figueroa JG, Borrás-Linares I, Lozano-Sánchez J, Segura-Carretero A (2018a) Comprehensive characterization of phenolic and other polar compounds in the seed and seed coat of avocado by HPLC-DAD-ESI-QTOF-MS. Food Res Int 105:752–763

Figueroa JG, Borrás-Linares I, Lozano-Sánchez J, Segura-Carretero A (2018b) Comprehensive identification of bioactive compounds of avocado peel by liquid chromatography coupled to ultra-high-definition accurate-mass Q-TOF. Food Chem 245:707–716

Furnier GR, Cummings MP, Clegg MT (1990) Evolution of the avocados as revealed by DNA restriction site variation. J Hered 81:183–188

Gamble J, Harker FR, Jaeger SR, White A, Bava C, Beresford M, Stubbings B, Wohlers M, Hofman PJ, Marques R, Woolf A (2010) The impact of dry matter, ripeness and internal defects on consumer perceptions of avocado quality and intentions to purchase. Postharvest Biol Technol 57:35–43

Ge Y, Cheng Z, Si X, Ma W, Tan L, Zang X, Wu B, Xu Z, Wang N, Zhou Z, Lin X, Dong X, Zhan R (2019a) Transcriptome profiling provides insight into the genes in carotenoid biosynthesis during the mesocarp and seed developmental stages of avocado (*Persea americana*). Int J Mol Sci 20(17):4117

Ge Y, Tan L, Wu B, Wang T, Zhang T, Chen H, Zou M, Ma F, Xu Z, Zhan R (2019b) Transcriptome sequencing of different avocado ecotypes: de novo transcriptome assembly, annotation, identification and validation of EST-SSR markers. Forests 10(5):411

Ge Y, Zhang T, Wu B, Tan BL, Ma F, Zou M, Chen H, Pei J, Liu Y, Chen Z, Xu Z, Wang T (2019c) Genome-wide assessment of avocado germplasm determined from specific length amplified fragment sequencing and transcriptomes: population structure, genetic diversity, identification, and application of race-specific markers. Genes 10(3):215

Gross-German E, Viruel MA (2013) Molecular characterization of avocado germplasm with a new set of SSR and EST-SSR markers: genetic diversity, population structure, and identification of race-specific markers in a group of cultivated genotypes. Tree Genet Genomes 9(2):539–555

Guzmán LF, Machida-Hirano R, Borrayo E, Cortés-Cruz M, Espíndola-Barquera MDC, Heredia-Garcia E (2017) Genetic structure and selection of a core collection for long term conservation of avocado in Mexico. Front Plant Sci 8:243

Hakizimana C, May J (2018) Can smallholder avocado production reduce poverty and improve food security through internal markets? The case of Giheta, Burundi. For Trees Livelihoods 27 (4):203–216

Hernández I, Uarrota V, Fuentealba C, Paredes D, Defilippi BG, Campos-Vargas R, Nuñez G, Carrera E, Meneses C, Hertog M, Pedreschi R (2022) Transcriptome and hormone analyses reveals differences in physiological age of 'Hass' avocado fruit. Postharvest Biol Technol 185:111806

Hurtado-Fernández E, Carrasco-Pancorbo A, Fernández-Gutiérrez A (2011a) Profiling LC-DAD-ESI-TOF MS method for the determination of phenolic metabolites from avocado (*Persea americana*). J Agric Food Chem 59(6):2255–2267

Hurtado-Fernández E, Pacchiarotta T, Gómez-Romero M, Schoenmaker B, Derks R, Deelder AM, Mayboroda OA, Carrasco-Pancorbo A, Fernández-Gutiérrez A (2011b) Ultra high performance liquid chromatography-time of flight mass spectrometry for analysis of avocado fruit metabolites: method evaluation and applicability to the analysis of ripening degrees. J Chromatogr A 1218(42):7723–7738

Hurtado-Fernandez E, Pacchiarotta T, Mayboroda OA, Fernandez-Gutierrez A, Carrasco-Pancorbo A (2015) Metabolomic analysis of avocado fruits by GC-APCI-TOF MS: effects of ripening degrees and fruit varieties. Anal Bioanal Chem 407(2):547–555

Hurtado-Fernández E, González-Fernández JJ, Hormaza JI, Bajoub A, Fernández-Gutiérrez A, Carrasco-Pancorbo A (2016) Targeted LC-MS approach to study the evolution over the harvesting season of six important metabolites in fruits from different avocado cultivars. Food Anal Methods 9:3479–3491

Ibarra-Laclette E, Méndez-Bravo A, Pérez-Torres CA, Albert VA, Mockaitis K, Kilaru A, López-Gómez R, Cervantes-Luevano JI, Herrera-Estrella L (2015) Deep sequencing of the Mexican avocado transcriptome, an ancient angiosperm with a high content of fatty acids. BMC Genomics 16(1):1–18

Jaganathan D, Bohra A, Thudi M, Varshney RK (2020) Fine mapping and gene cloning in the post-NGS era: advances and prospects. Theor Appl Genet 133(5):1791–1810

Juma I, Geleta M, Hovmalm HP, Nyomora A, Saripella GV, Carlsson AS, Fatih M, Ortiz R (2021) Comparison of morphological and genetic characteristics of avocados grown in Tanzania. Genes 12(1):63

Kämper W, Trueman SJ, Cooke J, Kasinadhuni N, Bruton AJ, Ogbourne SM (2021) Single-nucleotide polymorphisms that uniquely identify cultivars of avocado (*Persea americana*). Appl Plant Sci 9(6):e11440

Kilaru A, Cao X, Dabbs PB, Sung HJ, Rahman MM, Thrower N, Zynda G, Podicheti R, Ibarra-Laclette E, Herrera-Estrella L, Mockaitis K, Ohlrogge JB (2015) Oil biosynthesis in a basal angiosperm: transcriptome analysis of *Persea americana* mesocarp. BMC Plant Biol 15:203

Knight RJ Jr (2002) History, distribution and uses. In: Whiley AW, Schaffer B, Wolstenholme BN (eds) The Avocado: botany, production and uses. CABI Publishing, Wallingford, UK, pp 15–37

Kopp L (1966) A taxonomic revision of the genus *Persea* in the western hemisphere (Persea-Lauraceae). Mem N Y Bot Gard 14:1–120

Kuhn DN, Groh A, Rahaman J, Freeman B, Arpaia ML, Van den Berg N, Abeysekara N, Manosalva P, Chambers AH (2019a) Creation of an avocado unambiguous genotype SNP database for germplasm curation and as an aid to breeders. Tree Genet Genomes 15:71

Kuhn DN, Livingstone DS III, Richards JH, Manosalva P, Van den Berg N, Chambers AH (2019b) Application of genomic tools to avocado (*Persea americana*) breeding: SNP discovery for genotyping and germplasm characterization. Sci Hortic 246:1–11

Lahav E, Lavi U (2013) Genetics & Breeding. In: Schaffer B, Wolstenholme BN, Whiley AW (eds) The avocado: botany, production, and uses. CAB International 2013

Landahl S, Meyer MD, Terry LA (2009) Spatial and temporal analysis of textural and biochemical changes of imported avocado cv. Hass during fruit ripening. J Agric Food Chem 57:7039–7047

Lavi U, Hillel J, Vainstein A (1991) Application of DNA fingerprints for identification and genetic analysis of avocado. J Am Soc Hortic Sci 116:1078–1081

Le Nguyen K, Grondin A, Courtois B, Gantet P (2019) Next-generation sequencing accelerates crop gene discovery. Trends Plant Sci 24(3):263–274

Lee SK (1981) A review and background of the avocado maturity standard. California Avocado Society Yearbook 65:101–109

Lee SK, Young RE, Schiffman PM, Coggins CW (1983) Maturity studies of avocado fruit based on picking dates and dry-weight. J Am Soc Hortic Sci 108:390–394

Liu L, Shu B, Jue D, Wang Y, Wei Y, Shi S (2018) Avocado fruit pulp transcriptomes in the after-ripening process. Not Bot Horti Agrobot Cluj Napoca 47(2):308–319

López-Cobo A, Gómez-Caravaca AM, Pasini F, Caboni MF, Segura-Carretero A, Fernández-Gutiérrez A (2016) HPLC-DAD-ESI-QTOF-MS and HPLC-FLD-MS as valuable tools for the determination of phenolic and other polar compounds in the edible part and by-products of avocado. LWT Food Sci Technol 73:505–513

Lopez-Terrada MJ (2011) Flora and the Hapsburg crown: Clusius, Spain and American natural history. In: Dupré S, Lüthy CH (eds) Silent messengers: the circulation of material objects of knowledge in the early modern low countries. LIT Verlag, Berlin, pp 43–86

Lu Q-Y, Zhang Y, Wang Y, Lee R-P, Gao K, Byrns R, Heber D (2009) California Hass avocado: profiling of carotenoids, tocopherols, fatty acids, and fat content during maturation and from different growing areas. J Agric Food Chem 57:10408–10413

Marlett JS, Cheung T-F (1997) Database and quick methods of assessing typical dietary fiber intakes using data for 228 commonly consumed foods. J Am Diet Assoc 97:1139–1148

Mhameed S, Sharon D, Hillel J, Lahav E, Kaufman D, Lavi U (1996) Level of heterozygosity and mode of inheritance of variable number of tandem repeat loci in avocado. J Am Soc Hortic Sci 121(5):768–772

Miller P (1754) The gardener's dictionary, 4th edn, London

Pedreschi R, Ponce E, Hernández I, Fuentealba C, Urbina A, González-Fernández JJ, Hormaza JI, Campos D, Chirinos R, Aguayo E (2022) Short vs long-distance avocado supply chains: life cycle assessment impact associated to transport and effect of fruit origin and supply conditions chain on primary and secondary metabolites. Foods 11(12):1807

Perea Arango I, Padron IS, Litz RE (2010) A comparison of biotech methods for controlling infections by the avocado sunblotch viroid (ASBVd). In: Hortgen symposium, international horticultural congress, vol II, Lisbon. p. 280 (Abstract)

Pérez-Torres CA, Ibarra-Laclette E, Hernández-Domínguez EE, Rodríguez-Haas B, Pérez-Lira AJ, Villafán E, Alonso-Sánchez A, García-Ávila CJ, Ramírez-Pool JA, Sánchez-Rangel D (2021) Molecular evidence of the avocado defense response to *fusarium kuroshium* infection: a deep transcriptome analysis using RNA-Seq. PeerJ 9:e11215

Pliego-Alfaro F, Palomo-Ríos E, Mercado JA, Pliego C, Barceló-Muñoz A, López-Gómez R, Hormaza JI, Litz RE (2020) Chapter 12.1. *Persea americana* Avocado. In: Litz RE, Pliego F, Hormaza JI (eds) Biotechnology of fruit and nut crops, 2nd edn. CABI, pp 258–281

Popenoe W, Zentmyer GA, Schieber E (1997) The avocado has many names. Calif Avocado Soc 81:155–162

Raharjo S, Gomez-Lim MA, Padilla G, Litz RE (2008) Recovery of avocado (*Persea americana* mill.) plants transformed with the antifungal plant defensin gene PDF1.2. In Vitro Cell Dev Biol 44:254–262

Rendón-Anaya M, Ibarra-Laclette E, Méndez-Bravo A, Lan T, Zheng C, Carretero-Paulet L, Perez-Torres CA, Chacón-López A, Hernandez-Guzmán G, Chang T, Farr KM, Barbazuk WB, Chamala S, Mutwil M, Shivhare D, Alvarez-Ponce D, Mitter N, Hayward A, Fletcher S, Rozas J, Sánchez-Gracia A, Kuhn D, Barrientos-Priego AF, Salojärvi J, Librado P, Sankoff D, Herrera-Estrella A, Albert VA, Herrera-Estrella L (2019) The avocado genome informs deep angiosperm phylogeny, highlights introgressive hybridization, and reveals pathogen-influenced gene space adaptation. Proc Natl Acad Sci 116(34):17081–17089

Rivera SA, Ferreyra R, Robledo P, Selles G, Arpaia ML, Saavedra J, Defilippi BG (2017) Identification of preharvest factors determining postharvest ripening behaviors in 'Hass' avocado under long term storage. Sci Hortic 216:29–37

Rodríguez-López, C.E, C. Hernández-Brenes, V. Treviño, R.I. Díaz de la Garza (2017) Avocado fruit maturation and ripening: dynamics of aliphatic acetogenins and lipidomic profiles from mesocarp, idioblasts and seed. BMC Plant Biol 29:17(1):159

Rubinstein M, Eshed R, Rozen A, Zviran T, Kuhn DN, Irihimovitch V, Sherman A, Ophir R (2019) Genetic diversity of avocado (*Persea americana* mill.) germplasm using pooled sequencing. BMC Genomics 20(1):1–10

Ruiz-Chután JA, Berdúo-Sandoval JE, Kalousová M, Fernández E, Žiarovská J, Sánchéz-Pérez A, Lojka B (2022) SSRs markers reveal high genetic diversity and limited differentiation among populations of native Guatemalan avocado. J Microbiol Biotechnol Food Sci. https://doi.org/10.55251/jmbfs.6134

Salazar MJ, El Hafidi M, Pastelin G, Ramírez-Ortega MC, Sánchez-Mendoza MA (2005) Effect of an avocado oil-rich diet over an angiotensin II-induced blood pressure response. J Ethnopharmacol 98:335–338

Sánchez-Quezada V, Campos-Vega R, Loarca-Piña G (2021) Prediction of the physicochemical and nutraceutical characteristics of 'Hass' avocado seeds by correlating the physicochemical avocado fruit properties according to their ripening state. Plant Foods Hum Nutr 76(3):311–318

dos Santos MAZ, Alicieo TVR, Pereira CMP, Ramis-Ramos G, Mendonça CRB (2014) Profile of bioactive compounds in avocado pulp oil: influence of dehydration temperature and extraction method. J Am Oil Chem Soc 91:19–27

Schaffer B, Wolstenholme BN, Wiley AW (2013) Introduction. In: Schaffer B, Wolstenholme BN, Whiley AW (eds) The avocado: botany, production, and uses. CABI, Wallingford, pp 1–9

Scheben A, Batley J, Edwards D (2017) Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. Plant Biotechnol J 15(2):149–161

Schnell RJ, Brown JS, Olano CT, Power EJ, Krol CA, Kuhn DN, Motamayor JC (2003) Evaluation of avocado germplasm using microsatellite markers. J Am Soc Hortic Sci 128(6):881–889

Schroeder CA (1977) The persea tree of Egypt. California Avocado Society Yearbook 61:59–63

Serrano-García I, Hurtado-Fernández E, Gonzalez-Fernandez JJ, Hormaza JI, Pedreschi R, Reboredo-Rodríguez P, Figueiredo-González M, Olmo-García L, Carrasco-Pancorbo A (2022) Prolonged on-tree maturation vs. cold storage of Hass avocado fruit: changes in metabolites of bioactive interest at edible ripeness. Food Chem 394:133447

Sharma P, Al-Dossary O, Alsubaie B, Al-Mssallem I, Nath O, Mitter N, Rodrigues Alves Margarido G, Topp B, Murigneux V, Kharabian Masouleh A et al (2021) Improvements in the sequencing and assembly of plant genomes. Gigabyte 2021:1–10

Sharon D, Cregan PB, Mhameed S, Kusharska M, Hillel J, Lahav E, Lavi U (1997) An integrated genetic linkage map of avocado. Theor Appl Genet 95(5):911–921

Slater G, Shankman S, Shepherd JS, Alfin-Slater RB (1975) Seasonal variation in the composition of California avocados. J Agric Food Chem 23:468–474

Stahl AL (1933) Avocado maturity studies. Proc Fla State Hort Soc 46:123–133

Talavera A, Soorni A, Bombarely A, Matas AJ, Hormaza JI (2019) Genome-wide SNP discovery and genomic characterization in avocado (*Persea americana* mill.). Sci Rep 9(1):1–13

Torres AM, Bergh BO (1980) Fruit and leaf isozymes as genetic markers in avocado. J Am Soc Hortic Sci 105:614–619

Uarrota VG, Fuentealba C, Hernández I, Defilippi-Bruzzone B, Meneses C, Campos-Vargas R, Lurie S, Hertog M, Carpentier S, Poblete-Echeverría C, Pedreschi R (2019) Integration of proteomics and metabolomics data of early and middle season Hass avocados under heat treatment. Food Chem 289:512–521

USDA (U.S. Department of Agriculture) (2018). https://fdc.nal.usda.gov/fdc-app.html#/food-details/171706/nutrients

Vergara-Pulgar C, Rothkegel K, González-Agüero M, Pedreschi R, Campos-Vargas R, Defilippi BG, Meneses C (2019) De novo assembly of *Persea americana* cv. 'Hass' transcriptome during fruit development. BMC Genomics 20(1):1–14

van Zonneveld M, Larranaga N, Blonder B, Coradin L, Hormaza JI, Hunter D (2018) Human diets drive range expansion of megafauna-dispersed fruit species. Proc Natl Acad Sci USA 115:3326–3331

Melon Nutraceutomics and Breeding

Prashant Kaushik

Contents

Abstract

Nowadays, people are preferring consumption of fresh fruits and vegetables with delicious taste. Further fruits and vegetables have high nutritional values, since they are enriched in minerals, vitamins, protein, fiber, and some other beneficial components for human health. Compared with other fruits, melon has more antioxidant properties and is beneficial for human health. Melons are considered a good source of vitamins C, B6, and K, and a maximum of potassium and copper is found in the pulps of melon, after using the fruits. Melon seeds are dried and powdered to be consumed to get maximum health benefits. In addition, there are many phytochemicals and useful components in melons, which were identified to be antifungal, antibacterial, antiviral, and anti-inflammatory. These compounds present in the melons promote health in humans by maintaining good source of

P. Kaushik (✉)
Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Valencia, Spain
e-mail: prakau@doctor.upv.es

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_41

vitamins in the human body. To achieve the desired yield and enriched beneficial nutritional compounds for human health, conventional breeding can be started, but it takes too long time, and so to overcome the challenges of conventional breeding, a new approach, namely, molecular breeding has been started. In molecular strategy, the approaches, namely, marker assisted selection MAS, quantitative trait analysis, next-generation sequencing, genome-wide association mapping, genomic selection, high-throughput analysis, and structural and functional genomics have been started recently.

Keywords

Melon · Nutritional value · Nutraceutomics · Conventional breeding · Homologous recombination and molecular breeding

1 Introduction

Melons are unique yet important fruits consumed around the world. These fruits are known for their dietary fiber, vitamins, and minerals and are a good source of water. Melons are also known as summer fruits or summer melons. The fruits of these melons are referred to as pepo, a kind of berry. The melons belong to the Cucurbitaceae family and are sweet and fleshy, making them easy for consumption. Melons are native to Africa and are present in the southwest valleys of Asia, especially India and Iran. Later, the importance of the fruit was understood by the European countries and they spread to almost all parts of the world (Manchali et al. 2021).

The name melon is derived from the Latin word "melo pepo." The melons are used and adapted in a wide range of conditions with maximum diversity all around the world. The production of melons is topped by China, followed by Turkey, and India occupies the fourth place, behind the USA. In India, especially Uttar Pradesh and then Andhra Pradesh and Tamil Nadu are known to have the maximum area under melon production. Also, in the USA, California contributes to a maximum of 58% of total melon production, followed by Arizona, which contributes around 28% of the total melon production. Not only the aforementioned countries, but also many more countries cultivate melons, and increase of their production is due to their numerous beneficial and useful properties in promoting human health (Fraga et al. 2019).

There are numerous benefits of melon cultivation, which can provide a good source of income for farmers who begin melon cultivation. This crop needs a maximum of 85–95 days from sowing to harvest and is fully packed with nutritional benefits. Therefore, melon cultivation, which has been popular in recent years, has become highly important in the agriculture field under proper supervision and care (Sharma et al. 2020).

There are many species of melons with varied genome sizes, among which cantaloupe, watermelon, and honey dew are the best-known melons consumed in almost all parts of the world due to their sweet taste and flavor, along with added

health benefits. As stated, melons contain many useful properties, providing good health upon consumption. Melons are considered a good source of vitamin C, B6, and K. Also, a maximum of potassium and copper are found in the pulps of melons. Along with this, the seeds of these melons are dried and consumed or powdered and consumed to get maximum health benefits. Most importantly, these melons contain maximum water content; hence, they are called summer fruits, making them profitable to farmers for production in summer and consumption for added health benefits (Gómez-García et al. 2020). Not only the fruits but also the seeds of the melons are highly healthy, containing many benefits. The seeds of these melons contain high vitamin B, minerals, and fats and are a good source of dietary fiber. These seeds are also known to regulate blood pressure, boosting energy levels while keeping the skin healthy and glowing. Apart from these, the melon fruits are known to cool the body after consuming them as a juice, whereas the seeds keep the human body active with their nourishing properties. Some recent studies revealed that watermelons contain lycopene, which is a phytochemical compound having antioxidant and anti-inflammatory properties, keeping chronic diseases at bay (Fraga et al. 2019).

Proper diet, physical health, and a calm mind are always recommendable to promote health and to fix many internal health issues. With their nutritional properties, these melons were known to promote mental calmness and, more importantly, to cure malnutrition in humans and children. A healthy diet will always fix malnutrition problems by providing the body with all the needed elements for growth. Along with starchy grains, dairy, other fruits, and vegetables, melons are also recommended to be added to the diet to keep many health issues at bay. It is recommended that a person should consume at least five fruits per day, which should include melons, as a must. Due to the lower carbs and total fats in the fruit pulp, melons can be consumed for dieting and regulating weight (Sánchez et al. 2021). Also, the lower sodium content is good for promoting proper brain functioning and is good for maintaining blood pressure. Along with this, the seeds of melons are known to contain maximum health benefits, especially in lowering the cholesterol in the blood vessels, thereby promoting a healthy lifestyle. Due to the presence of folate in these melons, they are considered good for bone health and bone mineral density. Also, the presence of lutein and zeaxanthin in the melons makes them good for eyesight and vision. The above factors, i.e., water content, dietary fiber, vitamins, minerals, especially vitamin C, less sodium, lower cholesterol, and improving bone density and vision, make melon a wonderful crop with all the daily required nutrients useful for a healthy human being and protect one from the symptoms of malnutrition (Thakur et al. 2019).

Also, certain melon species like watermelons, honey dew, and cantaloupes are very healthy in promoting a wonderful diet and keeping away severe maladies. The phytochemicals and bioactive compounds, especially in watermelons, help in improving health by increasing the availability of antioxidants to the body. The lycopene, present not only in watermelons but also in tomatoes and pink guava, is a potential antioxidant that helps in reducing the growth and development of cancerous cells. Therefore, the presence of lycopene in watermelons has become the limelight in studying its antioxidant properties, making this species of melon very useful in arresting chronic diseases like cancer (Rolbiecki et al. 2021).

There are many conventional breeding methods in melons, which include intra-specific crosses that generate variations; pedigree methods, which include selection of two parents and producing hybrids with diverse characteristics, etc., which were not very successful in melons, as these hybridization methods have not changed for many years, and similar procedures in all the crops may not be successful always. The breeding procedures can be carried out only when the plants are sexually compatible, and the insertion of new traits is also possible if the plants are sexually mated with each other. Also, a major limitation in conventional breeding is that during crossing of the plants, many unwanted genes are also transferred along with the desirable traits showing reduction in yields. Therefore, the use of these methods needs to be supplemented and new methods need to be employed in order to achieve proper growth, yields, and nutritional qualities in plants (Sultana et al. 2014).

Therefore, the use of next-generation methods like high-throughput genotyping will improve the nature of the crop and yields as well. Certain gene editing techniques help in improving allelic variations in plants. Also, automated throughput phenotyping and bioinformatic tools help in improving the genotypes of the plants. The genome-wide single nucleotide polymorphism (SNP) discovery in nonmodel organisms also helps in developing suitable melon species that are highly beneficial to farmers and breeders. The above newly emerging next-generation methods help in understanding the genetic nature of the crops and promote the development of new varieties, thereby increasing the production and productivity of the crops (Yashiro et al. 2005).

In this chapter, we will describe the importance of melon and its many useful health properties in improving the immunity of humans and keeping chronic or harmful diseases at bay. The nature of conventional methods in how they affect the crops and the ability of next-generation methods to improve the production of the crops are also discussed, along with the genetic manipulations and the biotechnological tools to be implemented in order to increase the yields and the nutritional quality of melons.

2 Description on Nutritional Constituents

Melons were considered as useful sources of nutrition due to their properties and their nutritional constituents. The melons even though of different types contain varied nutritional parameters like minerals, vitamins, etc.

The nutritional components in the fruits and the seeds of melons include many important compounds useful for human health with the abundance of water (Tables 1 and 2). Not only the water content, melon fruits contain high amount of vitamin C which is around 36.5 mg followed by vitamin A, i.e., 169 μg. The most abundant minerals found in the fruits of melons are potassium and magnesium, i.e., 267 mg and 12 mg, respectively. Not only the fruits but also the dried seeds contain high content of nutrients like proteins, around 27.40%, followed by useful fiber, around 25% adding up the importance of melon seeds along with the fruit. Therefore, the nutritional components in the seeds and the fruits of different melons make them good for consumption as they are packed with a whole lot of nutrition useful to human body on a daily basis. Due to the presence of these important compounds in melons, they were recommended for intake on a daily basis for extreme health benefits.

Table 1 Nutritional constituents in melon fruits

Components	Quantities in 100 g of raw fruit
Water	90.5 g
Proteins	0.84 g
Carbohydrates	8.15 g
Total fats	0.18 g
Dietary fiber	0.8 g
Sugars	0.9 g
Vitamin C	36.5 mg
Vitamin A	169 μg
Potassium	267 mg
Magnesium	12 mg

Table 2 Nutritional constituents in melon seeds

Components	Quantities in raw fruit
Moisture	7.15%
Oil	30.55%
Proteins	27.40%
Ash	4.81%
Carbohydrates	29%
Fibers	25%
Phenolic compounds	Less quantities

Compared to all the other biochemical and nutrient molecules present in the melon fruits, glucose and fructose were the most abundant, which are converted into easily available sugar sucrose. These sugars or the biochemicals help in maintaining the metabolism of the plants and provide food for the growth of the fruits. The sucrose content directly corresponds to the sweet taste of the fruits, which is a major parameter for increasing the market value of fruits (Kolayli et al. 2010). Therefore, the production of glucose to carry out the plant's physiological functions is given in a flow chart in Fig. 1.

The available sucrose present in the plant system is utilized along with raffinose and stachyose, which is formed into fructose, glucose, and galactose that are considered as major biochemical compounds in melons and are again synthesized into simple sugars like sucrose, which is made available to the plants and also known for enhancing the sweet taste of the fruits.

3 Medicinal Properties in Relation to Human Health

There are many phytochemicals and useful components in melons, which were identified to be antifungal, antibacterial, antiviral, and anti-inflammatory. These compounds present in the melons promote health in humans maintaining good source of vitamins in the human body. Certain studies revealed that these medicinal

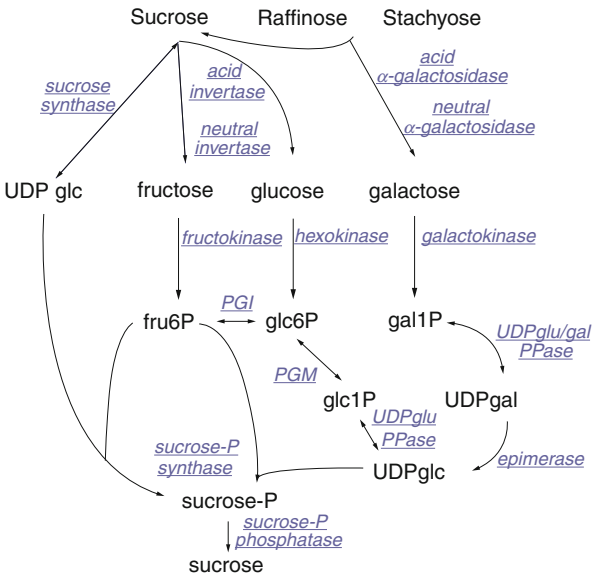


Fig. 1 A schematic diagram of the process of glucose production in plants

properties in the melons were useful in reducing many chronic diseases or the entry of such disease-causing agents can be arrested and killed immediately after it enters human body (Rajasree et al. 2016).

Melons provide maximum water content to the human body as they are abundant in water maintaining the cellular osmosis levels and promoting osmotic balance in human body. Moreover, the fruits of melons include important nutrients such as lycopene and citrulline, which are present because of the high levels of vitamins A and C. The lycopene is well-known to impart the red color to certain species of melons. Therefore, these red melons available in the market are well-known to have major health properties like reducing heart problems, arresting muscle soreness and treating inflammation in the human body wherever needed.

Certain physicians and doctors recommended that the intake of melon seeds helps in solving kidney problems, urinary bladder issues, and lowers high blood pressures thereby improving human health. The lycopene, which is a well-known antioxidant, helps in scavenging the free radicals in human body caused by radiations. The lycopene increases the body's tolerance ability toward sicknesses and ill health. Therefore, considering the above factors melons are considered as important sources of nutrition for humans to maintain balanced body weight and health (Manchali et al. 2021).

4 Genetic Resources of Health-Related (HR) Genes

The presence of particular genes in plants and the expressions of those genes have been shown to have a beneficial effect on human health. The presence of these genes in plants and the expressions of those genes have been shown to increase the

nutritional qualities of the plant, which in turn benefits the consumer. There were certain evidences of the presence of genes whose expressions can trigger and create phytomedicine synthesis in melons. The secondary gene pool of melons contains closely related plant species and hybrids. These cultivars are used for crossing and are known to produce fertile offspring. The interchange or crossing between the cultivars of this gene pool with primary gene pool is considered difficult because such crosses lead to the production of weak, sterile hybrids. Therefore, recivering the chromosomes of better parent becomes difficult. Finally, the tertiary gene pool produces completely sterile hybrids upon cross with the primary gene pool, because this gene pool is distantly related to the primary gene pool. Some special techniques are needed to transfer genes from wild gene pool to other gene pools, especially the primary one. These techniques are embryo rescue, chromosomal doubling, and bridge crosses with plants from the secondary gene pool (Pavan et al. 2017).

Cucumis melo, commonly known as melon, is a diploid plant species with a wide range of phenotypic diversity and economic importance. Melon fruits contain various nutraceuticals, such as carotenoids, flavonoids, vitamin C, and antioxidants, that have beneficial effects on human health (Garcia-Mas et al. 2012). The genetic basis of nutraceutical content in melon is not fully understood, but some candidate genes have been identified by GWAS and transcriptome analysis. For example, a gene encoding phytoene synthase (PSY), which catalyzes the first step of carotenoid biosynthesis, was found to be associated with β-carotene content in melon flesh (Pandey et al. 2016). Another gene encoding 9-cis-epoxycarotenoid dioxygenase (NCED), which regulates abscisic acid (ABA) synthesis from carotenoids, was found to be differentially expressed between climacteric and nonclimacteric melons. ABA is a hormone that influences fruit ripening and sugar accumulation. Moreover, some gene pools or germplasm collections of melon have been reported to have higher nutraceutical content than others. For instance, a wild relative of melon from India (*Cucumis melo* subsp. *agrestis* var. *momordica*) showed higher levels of antioxidants and phenolics than cultivated varieties (Zhang et al. 2022). Similarly, a landrace of melon from China (*Cucumis melo* var. *makuwa*) showed higher levels of metal ions than other types. These gene pools may provide valuable resources for breeding new varieties with improved nutraceutical content (Lotti and Fernández-Silva 2019).

5 Brief on Genetic Diversity Analysis

Melons were known for the differences in their phenotypical characters. These characters differ from one species to another in shape, size, and color of the fruits. Regarding melons, many studies previously suggested the changes in the phenotypic characters show diversity in their morphological parameters. Certain diversified traits were identified in case of melon's sex expression, fruit color, shape, number of sutures, layer around the seeds, flesh color, and the number of placentas. In case of seeds, the melons were observed to have small seed size, less seed weight, sometimes seedless in some fruits like watermelons, etc. The presence of such traits differing from the wild relatives and land races was observed due to inter-crossing or gene transfer techniques. Therefore, new methodologies and techniques help in

understanding the diversification in plants which ultimately leads to the changes in the phenotypic nature of the plants where sometimes these diversifications lead to positive changes, but sometimes they may be negative (Sensoy et al. 2007).

For example, Al-Zaher et al. (2016) characterized 50 snake melon accessions from Palestine using 17 phenotypic characters and found significant variation among them. They also identified four distinct groups of snake melons based on their morphological traits. Similarly, Yetisir et al. (2006) evaluated 56 Turkish melon genotypes using 12 phenotypic characters and found high diversity among them. They also classified the Turkish melons into six groups according to their botanical varieties. Other studies have used molecular markers such as RAPD3, SNP4, and SSR5 to complement phenotypic data and reveal more details about the genetic relationships and population structure of melon collections (Sharma et al. 2014). Phenotypic characters are useful for assessing genetic diversity in *Cucumis melo*, but they may be influenced by environmental factors and human selection. Therefore, combining phenotypic and molecular data can provide a more comprehensive picture of the genetic variation and evolution of this crop (Kimura-Kawakami et al. 2021).

5.1 Relationship with Other Cultivated Species and Wild Relatives

These melons, which were originated in South Africa, were cultivated in different parts of the world. The changes in the names and appearance of the melons make them different from each other. These melons contain different cultivated species which differ from each other, i.e., cucumber differs from gherkins where both these fruits belong to the same family. Likewise, these melons which belong to Cucurbitaceae family are closely related to each other. It was also observed that the taste, shape, and sizes of each melon species differ from each other which does not imply that they belong to different families. Therefore, the melons which were cultivated now in one area have a close relationship in sharing genes, characters, and pedigree with the other cultivable species in the same family (Mliki et al. 2001).

As India started the cultivation of melons long back, it was observed that many unexplored land races and wild relatives of melons having several useful characters were still unused. The accessions of such wild relatives were kept in store for future purpose as these wild relatives of any plants were known to contain many useful characters improving the yields, resistance against stresses, and the quality of the fruits. Understanding hybrid cultivars is a complex process. These cultivars come from crosses between wild relatives and cultivable species. This process takes a lot of time and effort. However, we can do this more efficiently now. We use techniques like marker-assisted breeding and genetic engineering. Although, some of the wild relatives of melons were explored to having many useful characters like maximum yields and stress resistance, many were yet to be explored as the process was time consuming (Endl et al. 2018).

Melons belong to the Cucurbitaceae family and have several wild relatives within the same genus, such as *C. amarus*, *C. mucusospermus*, *C. colocynthis,* and *C. zeyheri* (Zhang et al. 2019). These wild species are genetically very diverse and can provide useful genes for improving cultivated melons in terms of pest and disease resistance, drought tolerance, fruit quality, and other traits (Lotti and Fernández-Silva 2019).

Plant breeders have successfully crossed cultivated melons with some of these wild species, such as *C. amarus*, to introduce beneficial alleles into the melon gene pool. However, there are also challenges and limitations in using wild relatives for melon improvement, such as reproductive barriers, linkage drag, and genetic diversity loss (Guo et al. 2019). A comprehensive genome variation map of 1175 melon accessions and nine related species was recently published, which revealed the domestication history, population structure, and genetic diversity of melons and their wild allies. This resource could help identify novel genes and QTLs from wild species that could enhance melon-breeding programs (Zhao et al. 2019).

5.2 Relationship with Geographical Distribution

Melons are believed to have originated in Central Eastern Africa, and they spread throughout the world, possibly due to animals, humans, or birds carrying them. Musk melons come under Cucurbitaceae family with differences in their species, shapes, and sizes (Kerje and Grum 2000).

Certain species of melons called winter melons, *Citrullus,* and water melon were known to have originated in Africa and distributed to all over the world. Whereas, horned melon which originated in Africa was immediately grown in Australia, New Zealand, and Chile. Also, cantaloupe melon was cultivated in Italy. Honey dew and honey melon cultivars were cultivated in some parts of China. Mirza melon was cultivated in central Asia. Sharilyn melons, Galia, and Northern American cantaloupes were cultivated in European countries. There were many other cultivars which were grown in different parts of the world. Therefore, the above distribution of different melons all around the world in different countries provides an insight regarding the positive relationship of melons with respect to geographical distributions (Pavan et al. 2017).

5.3 Extent of Genetic Diversity

Melons are diploid species with $2n = 2x = 24$ chromosomes believed to be originated mostly in Africa and certain parts of Asia. After a few years, the melons' domestication was observed in Egypt, Europe, Middle East, China, and Afghanistan, whereas some other species of melons were imported from one country to another. Even though Africa is known to be the center of origin for melons, China was considered as the maximum producer and exporter of different hybrid melons. On an

average, these melons contain nearly 800 species or more under cultivation in different parts of the world (Nuñez-Palenius et al. 2008).

Several diversified characters and forms of melons were observed in different areas which include size ranging between 5 and 20 kgs, flesh color – orange, pink, red, white, etc., rind color – green white, yellow, red, etc., and form – round, flat, elongated, etc., based on different variations observed among different melon species. The changes in the sexual nature of the species, sex expression, and productivity of melons differ from species to species where some species gave low quality compared to the others. Even though many plant-breeding techniques emerged in melon cultivation they were observed to be less successful compared to newly emerging marker-assisted breeding and genetic transformation methods. These methods were also known to increase the diversity among the species by developing new, tolerant, and high-yielding cultivars in all the species of melons. Therefore, it was observed that the presence of maximum species of melons with less yields and low quality previously, the changes in the techniques for improving the yield, and other parameters gave rise to highly improved cultivars which were able to tolerate any stress up to a certain level, and such species were adaptable in different environmental conditions which was purely attained by the developed techniques under molecular biology (Guliyev et al. 2018).

6 Biotechnological Intervention for Health-Related Traits

Even though melons contain high nutrients, their shelf life, pollination, and storage create many problems in reducing the market value of the fruits.

Many scientists conducted experiments in melons to modify the fruits to be beneficial for consumption, and a few were successful in improving the yields and storage life of the fruits. Therefore, certain melon species were genetically modified for their ethylene biosynthesis in order to reduce the ripeness of the fruits to increase the shelf life for longer storage. On the other hand, seedless watermelons were developed by crossing male and female flowers. Also, certain gene modification in watermelons increased the sucrose content ultimately the sweetness of the fruits and the lycopene content, i.e., the antioxidant nature of the fruits so that they can be beneficial for consumption (Ezura et al. 2000).

There are many other aspects through which the melon fruits can be improved and made beneficial not only for consumption but also to increase the market value of the fruits. Therefore, it was understood that the implementation of molecular biology and biotechnology techniques in the improvement of melon fruits for different aspects was proven to be beneficial and useful in meeting the nutritional requirements after consumption. Biotechnology makes crop improvement faster and more accurate. It's a good alternative to traditional breeding methods, which take a lot of time (Ahmar et al. 2020).

Cucumis melo is a highly diverse crop with many cultivars and landraces that differ in morphological and agronomic traits. Several studies have assessed the

genetic diversity of melon collections based on phenotypic characters, such as fruit shape, size, color, texture, flavor, and yield.

6.1 Molecular Mapping for Health-Related Traits in Melon

Molecular mapping for health-related traits in melon is a technique that uses molecular markers to locate genes or quantitative trait loci (QTLs) that affect traits such as fruit quality, disease resistance, antioxidant content, or seed size. Molecular markers are DNA sequences that can be detected by various methods such as polymerase chain reaction (PCR), restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR), or single nucleotide polymorphism (SNP). These markers can be used to construct genetic linkage maps that show the relative position and distance of the markers on each chromosome (Diaz et al. 2011). By comparing the linkage maps of different individuals or populations with different phenotypes for a trait of interest, it is possible to identify the regions that are associated with the trait variation (Kumar and Sharma 2018). These regions are called QTLs, and they may contain one or more genes that influence the trait expression (Zhang and Liu 2017). Molecular mapping can help to identify candidate genes or regions that control these traits and facilitate marker-assisted selection or gene editing for improving melon breeding. Marker-assisted selection is a method that uses molecular markers to select individuals with desirable genotypes for a trait without relying on phenotypic evaluation (Leng and Xu 2017). Gene editing is a technique that uses engineered nucleases such as CRISPR-Cas9 to introduce targeted mutations or modifications in specific genes or regions of interest. Molecular mapping can also reveal the genetic diversity, population structure, and linkage disequilibrium of melon germplasm and provide insights into the evolution and domestication of different melon types (Liu and Zhang 2020). Genetic diversity is the amount and distribution of genetic variation within and among populations. Population structure is the pattern of genetic differentiation among populations due to factors such as geographic isolation, migration, selection, or drift. Linkage disequilibrium is the nonrandom association of alleles at different loci due to physical linkage or historical recombination events. Evolution and domestication are processes that shape the genetic and phenotypic changes in organisms over time due to natural or artificial selection pressures (Pandey and Saxena 2013).

6.2 Molecular Breeding and Genomics for Health-Related Traits in Melon

Molecular breeding and genomics for health-related traits in melon is a research field that aims to improve melon varieties by using molecular tools and genomic information. Health-related traits include disease resistance, fruit quality, antioxidant content, seed size, and other traits that affect human health and nutrition (Liu et al. 2020).

Molecular tools include molecular markers, genetic linkage maps, quantitative trait loci (QTLs), candidate genes, genome editing, and gene expression analysis. Genomic information includes genome sequences, transcriptomes, proteomes, and metabolomes of different melon types (Zhang and Liu 2017). Molecular breeding and genomics can help to identify genetic variation, dissect trait inheritance, select desirable genotypes, and modify target genes or regions for improving melon breeding (Pandey and Saxena 2021).

Molecular breeding and genomics are powerful tools for improving health-related traits in melon (*Cucumis melo* L.), such as disease resistance, fruit quality, and nutritional value (Chikh-Rouhou et al. 2023). Melon is a highly diverse crop with a wide range of morphological and physiological characteristics, making it a model plant for studying various aspects of plant biology. Several genomic resources have been developed for melon, including genome assemblies, molecular markers, gene expression data, and genetic maps (Liu et al. 2022). These resources enable the identification and characterization of quantitative trait loci (QTLs) or candidate genes associated with health-related traits in melon. For example, QTLs for resistance to Fusarium wilt, powdery mildew, cucumber mosaic virus, and melon necrotic spot virus have been mapped and validated in different melon populations (Zhao et al. 2021). Similarly, QTLs for fruit quality traits such as sugar content, flesh color, aroma, and antioxidant activity have been identified and used for marker-assisted selection (MAS) in melon-breeding programs. Moreover, genomics allows the discovery of novel genetic diversity and functional variation in melon germplasm collections that can be exploited for enhancing health-related traits in melon (Gonzalo et al. 2005). In addition to conventional breeding methods such as hybridization and selection, molecular breeding techniques such as MAS, genomic selection (GS), gene pyramiding (GP), and genome editing (GE) can be applied to improve health-related traits in melon by increasing the efficiency

+and accuracy of selection (Boualem et al. 2016). Molecular breeding and genomics have thus contributed significantly to the advancement of knowledge and innovation in melon research and development.

7 Conclusions

In conclusion, melons are delicious fruits with high nutritional value that offer numerous health benefits. They are packed with minerals, vitamins, fiber, and beneficial compounds, making them an excellent choice for a healthy diet. Melons, such as cantaloupe, watermelon, and honeydew, contain vitamins C, B6, and K, as well as potassium and copper. The seeds of melons also provide health benefits, including regulating blood pressure and promoting healthy skin. Additionally, melons possess medicinal properties, such as antioxidant and anti-inflammatory effects. Through molecular breeding approaches, researchers aim to enhance the growth, yields, and nutritional qualities of melons. Incorporating melons into our daily diet can contribute to overall well-being and protect against malnutrition.

References

Ahmar S, Saeed S, Khan MHU et al (2020) A revolution toward gene-editing technology and its application to crop improvement. Theor Appl Genet 133:1021–1041. https://doi.org/10.1007/s00122-019-03478-9

Al-Zaher M, Al-Masri A, Abu-Elbaiteh A (2016) Morphological characterization of snake melon (Cucumis melo var. flexuosus) landraces from Palestine. Genet Resour Crop Evol 63:303–313. https://doi.org/10.1007/s10722-015-0317-3

Boualem A, Troadec C, Campagne S et al (2016) Genetics, genomics and breeding of novelty in *Cucurbita pepo* summer squash. Theor Appl Genet 129:1079–1096. https://doi.org/10.1007/s00122-016-2715-x

Chikh-Rouhou H, Abdedayem W, Solmaz I, Sari N, Garcés-Claver A (2023) Melon (*Cucumis melo* L.): genomics and breeding. In: Singh S, Sharma D, Sharma SK, Singh R (eds) Smart plant breeding for vegetable crops in post-genomics era. Singapore, Springer Nature, pp 25–52

Diaz A, Fergany M, Formisano G et al (2011) A consensus linkage map for molecular markers and quantitative trait loci associated with economically important traits in melon (*Cucumis melo* L.). BMC Plant Biol 11:111

Endl J, Achigan-Dako EG, Pandey AK et al (2018) Repeated domestication of melon (*Cucumis melo*) in Africa and Asia and a new close relative from India. Am J Bot 105:1662–1671. https://doi.org/10.1002/ajb2.1153

Ezura H, Yuhashi K-I, Yasuta T et al (2000) Effect of ethylene on agrobacterium tumefaciens-mediated gene transfer to melon. Theor Appl Genet 101:75–79. https://doi.org/10.1007/s001220051406

Fraga CG, Croft KD, Kennedy DO et al (2019) The effects of polyphenols and other bioactives on human health. Food Funct 10:514–528. https://doi.org/10.1039/c9fo90042a

Garcia-Mas J, Benjak A, Sanseverino W et al (2012) The genome of melon (*Cucumis melo* L.). Proc Natl Acad Sci U S A 109:11872–11877. https://doi.org/10.1073/pnas.1205415109

Gómez-García R, Campos DA, Aguilar CN et al (2020) Valorization of melon fruit (*Cucumis melo* L.) by-products: phytochemical and biofunctional properties with emphasis on recent trends and advances. Trends Food Sci Technol 99:507–519. https://doi.org/10.1016/j.tifs.2020.02.024

Gonzalo MJ, Oliver M, Garcia-Mas J et al (2005) Simple-sequence repeat markers used in merging linkage maps of melon (*Cucumis melo* L.). Theor Appl Genet 110:802–811. https://doi.org/10.1007/s00122-004-1903-5

Guliyev N, Sharifova S, Ojaghi J et al (2018) Genetic diversity among melon (*Cucumis melo* L.) accessions revealed by morphological traits and ISSR markers. Turk J Agric For 42:393–401. https://doi.org/10.3906/tar-1711-48

Guo S, Zhao S, Sun H et al (2019) The genomes of four tapetal-lineage species reveal tapetum degeneration mechanisms during microsporogenesis. Nat Commun 10:4312. https://doi.org/10.1038/s41467-019-12927-9

Kerje T, Grum M (2000) The origin of melon, *Cucumis melo*: a review of the literature. Theor Appl Genet 510:37–44. https://doi.org/10.1007/978-94-011-4245-5_5

Kimura-Kawakami M, Kato-Murayama M, Yoshida T, Fukino N, Matsumoto S (2021) Genome-wide SNP markers for genotypic and phenotypic diversity analysis in melon cultivars and their wild relatives. Front Plant Sci 12:689. https://doi.org/10.3389/fpls.2021.633143

Kolayli S, Kara M, Tezcan F, Erim FB, Sahin H, Ulusoy E, Aliyazicioglu R (2010) Comparative study of chemical and biochemical properties of different melon cultivars: standard, hybrid, and grafted melons. J Agric Food Chem 58:9764–9769. https://doi.org/10.1021/jf101711w

Kumar R, Sharma SK (2018) Genomics-assisted breeding – a revolutionary strategy for crop improvement. In: Kumar R, Sharma SK, Chakrabarty SK, Singh AK (eds) Advances in plant breeding strategies: agronomic, abiotic and biotic stress traits. Springer, Cham, pp 1–36. https://doi.org/10.1007/978-3-319-94277-1_1

Leng PF, Xu ML (2017) Genomics-assisted breeding – a revolutionary strategy for crop improvement. J Integr Agric 16:2810–2826. https://doi.org/10.1016/S2095-3119(17)61789-3

Liu S, Wang X, Wang H (2022) Genetics, genomics and breeding in melon. Front Plant Sci 9:1487. https://doi.org/10.3389/fpls.2018.01487

Liu SY, Zhang YD (2020) Molecular mapping for fruit-related traits, and joint identification of candidate genes controlling seed size using high-density SNP-based genetic map in melon (*Cucumis melo* L.). Sci Hortic 272:109593. https://doi.org/10.1016/j.scienta.2020.109593

Liu SY, Zhang YD, Xu Y, Zhang XF, Wang YH, Li XW, Guo SL (2020) Melon (*Cucumis melo* L.): genomics and breeding. In: Varshney RK, Thudi M, Roorkiwal M, Sharma M, Singh NP, Gaur PM, Chaturvedi SK, Singh BB, Nadaf HL, Srivastava RK (eds) Legumes for soil health and sustainable management. Springer, Singapore, pp 17–38. https://doi.org/10.1007/978-981-15-2734-0_2

Lotti C, Fernández-Silva I (2019) Sweet genes in melon and watermelon. Nat Genet 51(11):1567–1568. https://doi.org/10.1038/s41588-019-0529-1

Manchali S, Chidambara Murthy KN, Patil BS (2021) Nutritional composition and health benefits of various botanical types of melon (*Cucumis melo* L.). Plan Theory 10:1755. https://doi.org/10.3390/plants10091755

Mliki A, Staub JE, Zhangyong S, Ghorbel A (2001) Genetic diversity in melon (*Cucumis melo* L.): an evaluation of African germplasm. Genet Resour Crop Evol 48:587–597. https://doi.org/10.1023/A:1011221822001

Nuñez-Palenius HG, Gomez-Lim M, Ochoa-Alejo N, Grumet R, Lester G, Cantliffe DJ (2008) Melon fruits: genetic diversity, physiology, and biotechnology features. Crit Rev Biotechnol 28:13–55. https://doi.org/10.1080/07388550801891111

Pandey S, Singh A, Kumar V (2016) Genome-wide differentiation of various melon horticultural groups for use in GWAS for fruit firmness and construction of a high resolution genetic map. Front Plant Sci 7:1437. https://doi.org/10.3389/fpls.2016.01437

Pandey SK, Saxena SC (2013) Map-based molecular diversity, linkage disequilibrium and association mapping of fruit traits in cucumber. Mol Breed 32:303–314. https://doi.org/10.1007/s11032-013-9863-6

Pandey SK, Saxena SC (2021) Advances in melon (*Cucumis melo* L.) breeding: an update. Sci Hortic 284:110024. https://doi.org/10.1016/j.scienta.2021.110024

Pavan S, Marcotrigiano AR, Ciani E, Mazzeo R, Zonno V, Ruggieri V, Lotti C, Ricciardi L (2017) Genotyping-by-sequencing of a melon (*Cucumis melo* L.) germplasm collection from a secondary center of diversity highlights patterns of genetic variation and genomic features of different gene pools. BMC Genomics 18:10. https://doi.org/10.1186/s12864-016-3448-9

Rajasree RS, Sibi PI, Francis F, William H (2016) Phytochemicals of cucurbitaceae family—a review. Theor Appl Genet 8(1):113–123. https://doi.org/10.25258/phyto.v8i1.6542

Rolbiecki R, Rolbiecki S, Figas A, Jagosz B, Wichrowska D, Ptach W, Prus P, Sadan HA, Ferenc P-F, Stachowski P (2021) Effect of drip fertigation with nitrogen on yield and nutritive value of melon cultivated on a very light soil. Theor Appl Genet 11(5):934. https://doi.org/10.3390/agronomy11050934

Sánchez E, Pollock R, Elkner T, Butzler T, Di Gioia F (2021) Fruit yield and physicochemical quality evaluation of hybrid and grafted field-grown muskmelon in Pennsylvania. Theor Appl Genet 7(3):69. https://doi.org/10.3390/horticulturae7030069

Sensoy S, Büyükalaca S, Abak K (2007) Evaluation of genetic diversity in Turkish melons (*Cucumis melo* L.) based on phenotypic characters and RAPD markers. Theor Appl Genet 54:1351–1365. https://doi.org/10.1007/s10722-006-9187-3

Sharma IP, Kanta C, Dwivedi T, Rani R (2020) Indigenous agricultural practices: a supreme key to maintaining biodiversity. In: Goel R, Soni R, Suyal DC (eds) Microbiological advancements for higher altitude agro-ecosystems and sustainability. Springer, pp 91–112

Sharma P, Singhvi NR, Solanki IS (2014) Assessment of genetic diversity in *Cucumis melo* through RAPD analysis. Indian J Agric Sci 84:251–254

Sultana RS, Mahabubur R, Rubaiyat A (2014) Melon crops improvement through biotechnological techniques for the changing climatic conditions of the 21st century. Int J Genet Genom 2:30–41

Thakur H, Sharma S, Thakur M (2019) Recent trends in muskmelon (*Cucumis melo* L.) research: an overview. J Hortic Sci Biotechnol 94:533–547

Yashiro K, Iwata H, Akashi Y et al (2005) Genetic relationship among east and south Asian melon (*Cucumis melo* L.) revealed by AFLP analysis. Breed Sci 55:197–206

Yetisir H, Sari N, Yücel S (2006) Some physiological and growth responses of watermelon [*Citrullus lanatus* (Thunb.) Matsum. and Nakai] grafted onto different rootstocks to flooding. Theor Appl Genet 58:1–8

Zhang X, Guo S, Ren Y et al (2019) Improving watermelons by harvesting genes from wild species. Theor Appl Genet 195:3–12

Zhang X, Guo S, Ren Y (2022) Mining candidate genes related to heavy metals in mature melon (Cucumis melo L.) fruit and pericarp using RNA-Seq analysis. Theor Appl Genet 13:1767. https://doi.org/10.3390/genes13101767

Zhang XP, Liu CG (2017) Genetics, genomics, and breeding in melon. Crit Rev Plant Sci 36:427–447

Zhao G, Liu W, Zhang J (2021) Advances in melon (*Cucumis melo* L.) breeding: an update. Sci Hortic 287:110381

Zhao G, Wang Y, Liu W (2019) A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. Nat Genet 51:1578–1590. https://doi.org/10.1038/s41588-019-0522-8

Guava: A Nutraceutical-Rich Underutilized Fruit Crop

Malarvizhi Mathiazhagan, Vasugi Chinnaiyan, and Kundapura V. Ravishankar

Contents

Abstract

Psidium guajava L., commonly known as guava, is native to the tropical and subtropical regions. It is one of the important commercial fruit crops rich in nutrients and phytochemicals that have many medicinal and nutritional benefits. It is now being recognized as "super fruit" due to the attractive color and compounds, which are known to be active as neutralizers of free radicals and are beneficial to human health. Guava fruit, leaf, bark, and seeds contain bioactive compounds that can

M. Mathiazhagan · K. V. Ravishankar (✉)
Division of Biotechnology, ICAR–Indian Institute of Horticultural Research, Bengaluru, India
e-mail: ravishankar.kv@icar.gov.in

V. Chinnaiyan
Division of Fruit crops, ICAR–Indian Institute of Horticultural Research, Bengaluru, India

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_42

function as antioxidants and anti-inflammatory agents, thereby helping in alleviating different types of cancer, diabetes, and infectious diseases and promoting overall health. Guava genetic resources are diverse, with more than 400 germplasm and cultivars available around the world. Morphogenetic and molecular studies have improved our understanding of the guava germplasms. This chapter discusses in detail about the health benefits of phytochemicals, as well as genetic and molecular studies for improvement of health-related traits in guava.

Keywords

Guava · Phytochemicals · Bioactive compounds · Molecular markers · QTL · Nutraceutical

1 Introduction

Guava (*Psidium guajava* L) is widely grown across the tropical and subtropical countries of the world. Increased public awareness toward balanced diet along with nutrient benefits of guava has resulted in a global market demand for the fruit. In addition to the fruit, leaf, bark of stem, root, and seeds are also rich in phytochemicals, vitamins, and minerals. Because of their antioxidant, antidiabetic, anti-inflammatory, anticancerous, and antidiarrheal properties, many parts of the plant are found to be beneficial to human health (Shanthirasekaram et al. 2021; Blancas-Benitez et al. 2022) (Fig. 1). Further, because of its climacteric nature, the shelf-life of the fruit is less and

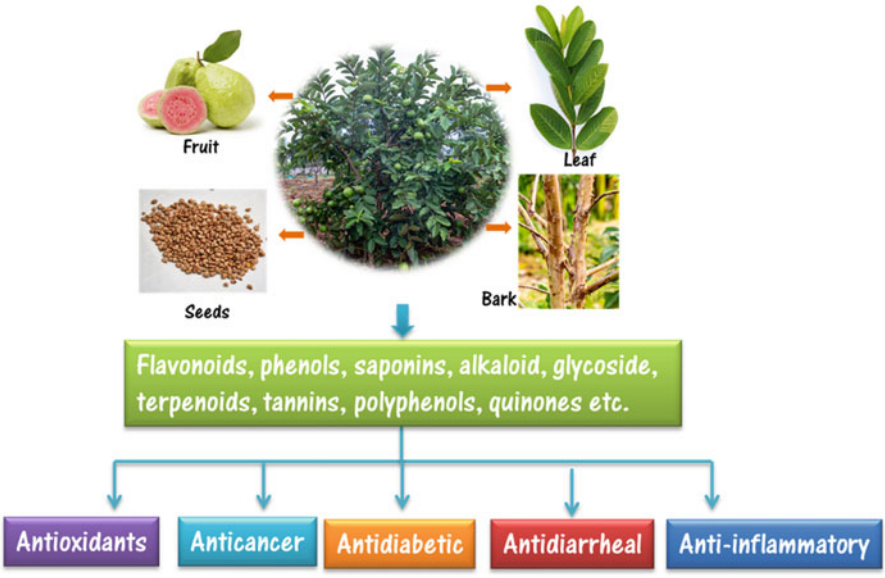


Fig. 1 Benefits of different parts of guava for human health

warrants immediate consumption. Hence, to mitigate this post-harvest loss, a variety of processed and functional food products like guava purees, jams, jellies, RTS beverages, ice creams, tea, etc. have been developed, which preserve the bioactive compounds found in the fruit and thereby catering the nutritional needs of the consumers (Sampath Kumar et al. 2021) have been developed by various workers. In order to exploit the genetic resources available in guava, morphological and genetic diversity studies play a major role. Different molecular markers have been employed to characterize and analyze the variability of guava accessions and cultivars (Pavani et al. 2022; Usman et al. 2020). Because it is an underutilized crop, there have been little efforts to understand the quantitative trait loci (QTLs) or genes controlling the important fruit quality and health-related traits. In addition, application of marker-assisted selection (MAS) and genome-assisted breeding approaches such as genome-wide association studies (GWAS), genomic selection (GS), and genome editing has not been understood in guava. With the advancements in sequencing technologies and metabolomics, efforts to identify health-related traits like vitamin C content, lycopene content, shelf-life, etc. are expected to accelerate in the future.

2 Phytochemicals with Nutraceutical Properties

Various studies on the physical, chemical, and sensory attributes of guava fruit, leaf, seed, and stem bark have been undertaken all over the world. The fruit is a rich source of vitamin C (which is four times higher than orange) and other vitamins like vitamin A (retinol), vitamin B_1 (thiamine), vitamin B_2 (riboflavin), vitamin B_3 (niacin), and vitamin B_6 (pyridoxine) (Kumari et al. 2018); minerals like calcium, magnesium, phosphorus, potassium, zinc, and iron; proteins; carbohydrates; dietary fiber (5.2 g/100 g); and lycopenes and carotenes (USDA 2019; Pandian and Jayalakshmi 2019). It's also a good source of pectin (0.5–1.9%). Many pharmacological studies have shown that the phytochemicals found in different parts of the guava plant possess various health benefits and have medicinal applications. The composition and concentration of several chemical compounds present in guava differ among the cultivars and are also influenced by the horticultural practices adapted.

2.1 Guava Leaf

The leaves are simple, opposite, oblong to elliptic with an entire margin, and have a specific aroma when crushed, which comes from an essential oil and smell depends on the cultivar. Because of its antioxidant potential and numerous phytochemicals which can boost health and ameliorate the effects of ailments or diseases, guava leaf tea is a better herbal tea alternative to commercial tea. The leaves have long been used in traditional medicine to treat diarrhea (Birdi et al. 2020) and other gastrointestinal ailments and infections. The leaf extracts have been used to treat a variety of pathogenic bacteria and viruses as natural antimicrobial agents. Some Ethiopian guava cultivars were found to have good antagonistic activity against *Salmonella*

typhi and *Shigella boydii* (Oncho et al. 2021), whereas various commercial guava cultivars with high phenols, flavonoids, and increased antioxidant activity were found to have a high level of inhibition against *Escherichia coli* and *Bacillus subtilis* (Melo et al. 2020). Guava leaf extracts are more effective against Gram positive than Gram negative bacteria in many studies (Abdullah et al. 2019). In addition, guava leaf tea has been demonstrated to be effective in suppressing influenza H1N1 viruses (Sriwilaijaroen et al. 2012). The antimicrobial activity of the leaf extract was found to differ based on the solvents used for extraction (Biswas et al. 2013). The multitude of phytochemicals such as flavonoids, phenols, saponin, alkaloid, glycoside, terpenoids, tannins, polyphenols, and quinones found predominantly in the guava leaf extracts were found contributing to its plethora of health benefits (Shanthirasekaram et al. 2021). In addition, guava leaf extracts can also act as natural immune stimulants, enhancing our immune system's ability to fight infections (Laily et al. 2015).

The mouthwash of leaves was effective for aphthous ulcers in terms of reduction of symptoms of pain and faster reduction of ulcer size (Guintu and Chua 2013). The presence of functional compounds or the bioactive poly-phenolic compounds like quercetin, caffeic, ferulic, ascorbic, and gallic acids acts as antioxidants in the treatment of numerous diseases (Denny et al. 2013). The leaf extract has been found to be promising as a herbal medicine to treat hyperglycemia. The leaves were shown to possess antidiabetic activity in diabetic mice, protective activity on liver cells (Zhu et al. 2020), reduced oxidative stress, inhibition of inflammation, and β-cell death (Jayachandran et al. 2018). The bioactive substances such as triterpenoids, monoterpenoids, flavonoids, and phenolic compounds have been discovered as important components in the regulation of type 2 diabetes mellitus (Jiang et al. 2021). These compounds control the insulin secretion and thereby lower the blood glucose level. Furthermore, due to their high free radical scavenging activity, polysaccharides from guava leaf have shown potential to be used as an antidiabetic or antioxidant compounds (Luo et al. 2019). Guava leaves also have anticancer activity and show growth inhibition of cancer cells and inducing apoptosis. Kampferol in guava leaves have been proven to have scavenging and antiproliferative activity on thyroid cell lines (Kim 2011), betulinic acid on human cholangiocarcinoma cell lines (Phonarknguen et al. 2022), polyphenols on breast, lung, and prostate cancer cell lines (Alhamdi 2019).

As an alternative, because of their antiproliferative, antioxidant, and antibacterial properties, essential oils from guava leaf have also been used for treating various diseases (Rakmai et al. 2018). The guava leaf oil is a rich source of volatiles like Caryophyllene, 1, 8-Cineole, Limonene, a-Pinene, and different proportions of terpenes and hydrocarbons (Soliman et al. 2016). As the leaves possess antimicrobial compounds, leaf extracts and essential oils are commonly used as a medicine against gastroenteritis.

2.2 Guava Fruit

The fruit is a berry with a thick pericarp and pulpy seed cavity, emitting a strong, sweet, musky odor when ripe. The shape ranges from glucose to sub-globose,

spherical, oblong, ovate, and pyriform, with four or five protruding floral remnants (sepals) at the apex weighing up to 500 g depending upon the cultivars. The fleshy mesocarp is of varying thickness and has a softer endocarp with numerous small, hard yellowish-cream seeds embedded throughout (Malo and Campbell 1994). The skin is thin, and the color of ripe fruit varies from pale yellow to dark yellow or blushed with pink, sometimes completely pale green or dark purple, depending upon the cultivars. The pulp color may be white, creamy white, creamy yellow, pale pink to dark red, purple to orange, slightly juicy, acid, sub-acid, or sweet and flavorful. The fruits are rich sources of carbohydrates (12.16%), proteins (2.3%), fats (0.7%), ascorbic acid (241.86 mg/100 g), fiber (4.8%), and minerals like calcium (17.63 mg), iron (0.24 mg), and zinc (0.21 mg/100 g), and moisture content (84.31%). The calorie yield is around 68 calories/100 g of fruit (Bogha et al. 2020). Among the pulp colors, the pink/red pulp types are more nutritious due to the presence of higher amount of phytochemicals and secondary metabolites like phenolic compounds, flavonoids, carotenoids, and tannins, which contribute to its antioxidant potential (Suwanwong and Boonpangrak 2021) than the white pulped fruits. The whole fruit is generally consumed, as the peel is rich in ascorbic acid and phenolic compounds compared to pulp (Musaa et al. 2015; Emanuel et al. 2018). The pink or red pulped guava fruits have shown to be a promising therapeutic alternative due to their increased antioxidant potential and high levels of lycopene and vitamin C (Nwaichi et al. 2015).

Numerous studies have emphasized the importance of using lycopene-rich guava fruit extract for treating different ailments, including reducing oxidative stress damage in human dermal fibroblasts (Alvarez-Suarez et al. 2018); ameliorative effects on cigarette smoke-damaged pulmonary tissues (Meles et al. 2021); cytostatic and cytotoxic effects on breast cancer cells (Dos Santos et al. 2018); modulating colon health (Blancas-Benitez et al. 2022); inhibition of intestinal resorption of glucose (König et al. 2019); frailty reduction and health promotion (Ruangsuriya et al. 2022). Pharmacological studies on pink-pulped guava fruits recorded increased hemoglobin levels in pregnant women (Sormin et al. 2020); antiplatelet activity (Rojas-Garbanzo et al. 2021); and apoptosis induction in breast cancer cells (Liu et al. 2020; Polinati et al. 2022). Guavinoside B, a biologically active compound found in guava fruit, displayed inhibitory activities on α-glucosidase, thereby providing a new therapeutic possibility for people with diabetes (Xu et al. 2022). Additionally, guava pulp, seed, and leaf phenols have shown to improve the diabetic parameters in albino rat model (Shabbir et al. 2020). Presence of high amounts of ascorbic acid, flavonoids, and polyphenols makes guava fruit a suitable candidate as a natural therapeutic agent.

2.3 Guava Seed

The seeds are good source of proteins, fats, dietary fiber, phenolics, flavonol, glycosides, tannins, saponins, and vitamins. They possess free radical scavenging activities due to the presence of phenolic compounds like chlorogenic acid, apigenin and its glycosides, caffeic acid derivatives, quercetin, phytosterols and fatty acids

like ethylpalmitate, linoleic acid, stigmasterol, and campesterol, and dietary fiber (Prommaban et al. 2020; Emanuel et al. 2018). Additionally, the polysaccharides from guava seeds were also found effective in immunomodulation and inhibition of breast cancer cell growth (Lin and Lin 2020) and prostate cancer cell growth (Lin and Lin 2021). Further, the seed oil has exhibited wound healing properties and also demonstrated growth inhibition of human erythroleukemic cells (Prommaban et al. 2019). In addition, the seed protein hydrolysates displayed α-amylase inhibition, thereby making it a suitable plant peptide-based antidiabetic agents (Jamesa et al. 2020). The seeds, generally obtained as a by-product of processing industry, have the potential to be utilized as a functional food, food additives, or structural modulators in addition to their nutraceutical applications. An elaborate review on the uses of guava seeds in food processing and nutraceutical applications is available for further reading (Kumar et al. 2022).

2.4 Guava Bark

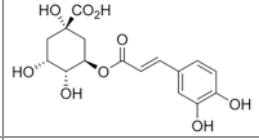
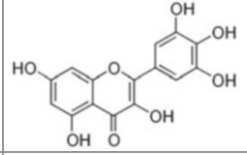
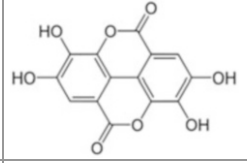
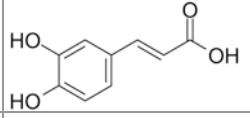
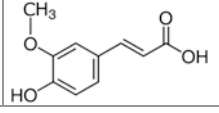
Along with the leaves, guava stem bark has been used in traditional medicine to treat diarrhea, dysentery, and fever. The presence of bioactive phenolic compounds like gallic acid, chlorogenic acid, caffeic acid, and ellagic acid and flavonoids like catechin, rutin, quercetin, and luteolin renders guava stem bark with good antimicrobial (Gurmachhan et al. 2020), antidiabetic, antinephrolithiatic (Irondi 2020), and other health benefits (Table 1). The presence of flavonoids, tannins, alkaloids, and saponins in the root barks of guava was also reported to have antimicrobial and antioxidant activities (Kuber et al. 2013).

2.5 Post-Harvest Techniques

The fruits are climacteric, ripens fast with increase in rate of respiration and metabolic activities due to which the shelf life of fruits does not last for more than 5–7 days and making them unpleasant and commercially undesirable. Hence, there is a need for processing raw fruit into ready-to-eat products with a longer shelf life coupled with enhanced nutritional attributes. It is processed into wide an array of products like juice, jam, jelly, puree, concentrate, cheese, toffee, fruit flakes, squash, syrup, nectar, powder, wine, and vinegar, as well as ready-to-eat snacks, drinks, ice cream, biscuits, dehydrated, and canned products (Kumari et al. 2018; Sampath Kumar et al. 2021; Kumar and Gupta 2021). The by-products obtained from pulp industries are generally used as food additives, pharmaceuticals, biofuels, etc. (Iha et al. 2018; Anjali and Manjul 2021; Hoyos et al. 2022) and serve as promising functional foods due to the presence of a large number of phytochemicals (da Silva Lima et al. 2019).

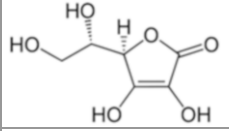
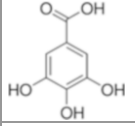
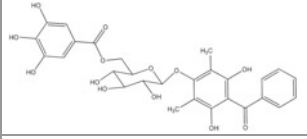
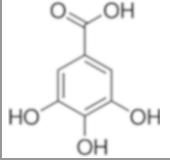
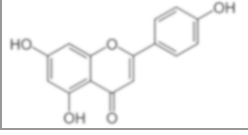
Guava leaves, fruits, and seeds have become functional foods due to their health benefits. Biscuits made from guava seeds and pulp have been reported to have a significant effect in reducing the blood glucose levels in diabetes patients

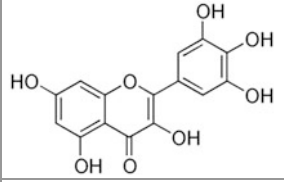
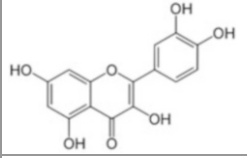
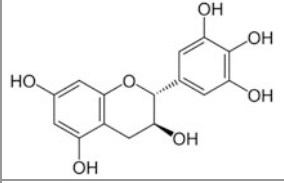
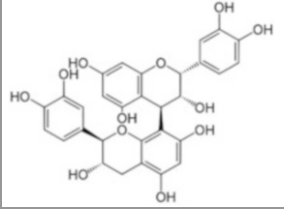
Table 1 Prominent bioactive compounds present in guava

Bioactive compounds	Structure	Plant source	Properties	Reference
Phenolic and poly-phenolic compounds				
Chlorogenic acid		Leaf, seed, bark	Antioxidant, antimicrobial	Prommaban et al. 2020; Gurmachhan et al. 2020
Myricetin		Fruit, seed	Antioxidant	Emanuel et al. 2018
Ellagic acid		Seed, bark	Antioxidant	Prommaban et al. 2020; Gurmachhan et al. 2020
Caffeic acid		Leaf, bark	Antioxidant	Denny et al. 2013; Gurmachhan et al. 2020
Ferulic acid		Leaf	Antioxidant	Denny et al. 2013

(continued)

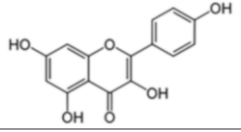
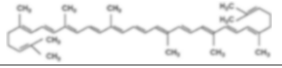
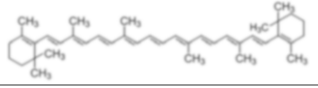
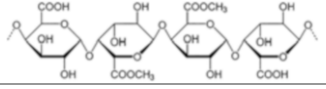
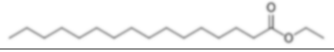
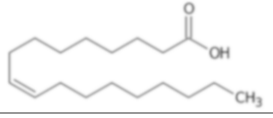
Table 1 (continued)

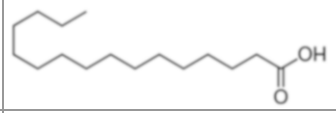
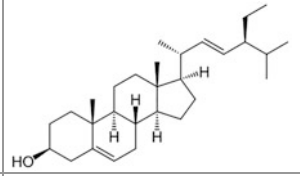
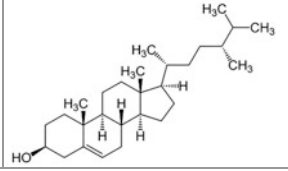
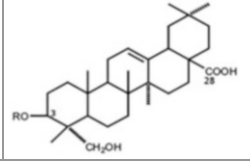
Bioactive compounds	Structure	Plant source	Properties	Reference
Ascorbic acid		Leaf, fruit	Antioxidant, anti-inflammatory	Suwanwong and Boonpangrak 2021
Gallic acid		Leaf, bark	Antioxidant	Melo et al. 2020; Gurmachhan et al. 2020
Guavinoside B		Fruit	Antidiabetic	Xu et al. 2022
Tannins		Leaf, bark	Antioxidant, antimicrobial	Oncho et al. 2021
Flavonoids				
Apigenin		Fruit, seed	Antioxidant and antimicrobial	Prommaban et al. 2020; Melo et al. 2020

Myrcetin		Fruit	Antioxidant and antimicrobial	Shukla et al. 2021
Quercetin		Leaf, bark	Antioxidant, antidiarrheal	Emanuel et al. 2018; Gurmachhan et al. 2020
Gallocatechin		Fruit, leaf	Antioxidant	Blancas-Benitez et al. 2022
Procyanidin B		Fruit	Antioxidant	Prommaban et al. 2020

(continued)

Table 1 (continued)

Bioactive compounds	Structure	Plant source	Properties	Reference
Kampferol		Leaf	Antioxidant, anticancer	Phonarknguen et al. 2022
Carotenoids				
Lycopene		Fruit	Antioxidant	Shukla et al. 2021
B-carotene		Fruit	Antioxidant	Nwaichi et al. 2015
Polysaccharides				
Pectin		Fruit	Antidiabetic and antioxidant	Chang et al. 2020
Fatty acids				
Ethyl palmitate		Seed	Antioxidant	Prommaban et al. 2020
Linoleic acid		Seed	Antioxidant	Prommaban et al. 2020
Oleic acid		Seed	Antioxidant	Prommaban et al. 2020

Palmitic acid		Seed	Antioxidant	Prommaban et al. 2020
Stigmasterol		Seed	Antioxidant	Prommaban et al. 2020
Campesterol		Seed	Antioxidant	Prommaban et al. 2020
Alkaloids				
Saponins		Leaf, bark	Antioxidant, antimicrobial	Oncho et al. 2021

(Kaushik 2019). Further, guava seeds have found applicability in anti-acne face wash gels that harnesses their antioxidant and antimicrobial properties (Kamble et al. 2019). Similarly, guava purees rich in vitamin C and polyphenols have been reported to have a positive effect on the hyperglycemic and hypercholesterolemic rats (Pérez-Beltrán et al. 2017). Guava acts as a prebiotic and improves the flavor and nutritional value of traditional dairy products like yogurt, whey beverages, mousse, and smoothies (Chauhan et al. 2015).

In addition to adding value, post-harvest packaging techniques are critical for maintaining the fruit quality and nutritional attributes for an extended period of time (Etemadipoor et al. 2020). Different packaging techniques like modified atmosphere packaging (MAP) (Mangaraj et al. 2021), controlled atmosphere packaging (CAP), edible packaging (Murmu and Mishra 2018), antimicrobial/antifungal packaging (Elabd 2018), and nano packaging (Kalia et al. 2021) are available to extend the shelf life with minimal changes to the physiochemical properties of the fruit. Overall, post-harvest management, value-addition and processing in guava has gained significance in recent years with different kinds of functional foods being introduced into the market. However, in order to preserve the fruit's functional properties in the processed goods, the guava processing technologies must be improved, and in-depth research is needed.

2.6 Biotechnological Interventions to Improve Nutraceutical Properties

Besides some phytochemical analysis and basic pharmacological studies, the information on the mechanism of action of bioactive compounds and their antioxidant pathways are scarce in guava. In order to speed up guava breeding efforts, research on molecular markers, genes, and QTLs linked to nutraceutical properties such as vitamin C, lycopene content, shelf life, etc. are essential. Further, gene expression studies, association mapping, prediction modeling (genomic selection), and genome editing techniques can help expedite the efforts in identifying and breeding new guava cultivars with enhanced nutritional and therapeutic applications.

3 Genetic Resources

Guava is an important commercial crop in the tropical and subtropical countries and is cultivated in India, Mexico, Brazil, Thailand, Spain, Portugal, Southern France, Israel, Panama, EI Salvador, Costa Rica, Nicaragua, Bolivia, Malaysia, Kenya, the USA (Hawaii, California, and Florida), New Zealand, the Philippines, China, Indonesia, Cuba, Java, Venezuela, Pakistan, Australia, and some African countries, with India contributing around half the world's guava production. From tropical American countries, guava reached Asian and African countries through its colonizers like Spain and Portugal. It is reported that there are more than 400 guava cultivars are available around the world, majority of which belongs to the cultivated species,

Psidium guajava L. There are other wild species that has a wide variability in fruit shape, size, and flavor. Some important species which are used in breeding program for imparting resistance to biotic and abiotic stresses are: Guinea guava or Castilian guava or Brazilian guava (*Psidium guineense* Sw syn: *P. Molle* Bertol, *P. Araca* Raddi, *P. Schiedeanum* O.Berg.), Strawberry or cherry guava or cattley guava or Chinese guava (*P. Cattleianum* Afzel. ex Sabine syn; *Psidium littorale* Raddi, *P. littorale* (Raddi) var. *longipes* (O.Berg.), *P. cattleianum* var. littorale (Raddi) Mattos, *P. cattleianum* var. Purpureum Mattos, *P. Cattleianum* var. Pyriformis Mattos), Costa Rican guava or Cas guava (*Psidium friedrichsthalianum* (O.Berg) Niedenzu, syn: *Calyptropsidium friedrichsthalianum*) (Rajan and Hudedamani 2019). Prominent guava cultivars around the world for commercial and breeding works are listed here (Table 2).

Guava fruit quality parameters like shape, size, TSS, acidity, peel color, pulp color, pulp recovery, pectin content, ascorbic acid, phenols, dietary fiber, and shelf life vary greatly among guava germplasm collections around the world. Some of the traditional and modern cultivars rich in health-related traits are listed in Table 3.

4 Genetics and Traditional Breeding

Information on the genetics and inheritance of genes related to commercial traits is limited due to the perennial nature and heterozygosity of the guava plant. Guava breeding programs have generated segregating populations/progenies by crossing contrasting cultivars to understand the segregation phenomenon governing important quantitative traits like fruit peel color, pulp color, seed characters, fruit weight, yield, TSS, acidity, leaf characters, and tree characters. But only a small portion of the genetic diversity available in the crop has been utilized for commercial purpose, and the good amount of diversity that exists in the wild progenitors needs to be explored.

When it comes to pulp color, red was reported to be dominant over white pulp, with a segregation ratio of 3:1. In addition, red pulp has shown strong linkage with large and hard seeds, whereas white pulp has been linked with small and soft seeds (Subramanyam and Iyer 1990). The fruit quality traits like TSS in guava appear to be due to non-additive gene effects, while fruit weight and seed hardiness exhibited a negative association at the phenotypic and genotypic levels (Dinesh et al. 2017). The characters like fruit yield, fruit weight, outer flesh thickness, acidity, number of seeds per fruit, seed weight, plant height, phenol, and pectin content were governed by additive gene action and had high heritability combined with high genetic advance as percent of mean (Patel et al. 2015; Singh et al. 2015a; Gupta and Kour 2019). Further, a few studies have reported a low genetic variability and a high phenotypic variability for characters like fruit mass, fruit width, sepal size, length of stalk, external flesh thickness, internal flesh thickness, and external/internal flesh thickness ratio. This indicates the influence of environmental conditions over the phenotypic variability of individual progenies in a population. In contrast, the length of fruit and the ratio of diameter of calyx/fruit had a very high heritability due to additive effects

Table 2 Prominent guava cultivars grown around the world

Country	Cultivar name	Reference
Australia	'Northern Gold', 'Thai White' ('Glom Sali'), 'Common', 'Cherry', 'Allahabad Safeda', 'Beaumont', 'Sardar', 'KaHua Kula'	Mitra and Thingreingam Irenaeus 2016; Rajan and Hudedamani 2019
Bangladesh	'Swarupkathi', 'Mukundapuri', 'Kanchannagar', 'Kazi'	Rajan and Hudedamani 2019
Brazil	Paluma', 'Rica', 'Pedro Sato', 'Seculo XXI' 'Sassaoka, 'Yamamoto', 'XXI Century', 'Kumagai'	Mitra and Thingreingam Irenaeus 2016; Rajan and Hudedamani 2019
China	'Pearl', 'Red Heart', 'China pear'	Tan et al. 2020; Jaleel et al. 2021
Cuba	EnanaRoja', 'Cubana', 'EEA 1-23', 'Supreme Roja'	Mitra and Thingreingam Irenaeus 2016
Egypt	'Bassateen El Sabahia', 'Bassateen Edfina', 'Allahabad Safeda'	Rajan and Hudedamani 2019
India	'Allahabad Safeda', 'Swetha', 'Dhawal', 'Apple Colour', 'Lucknow-42', 'Lucknow-49', 'Safeda', 'Seedless', 'Red Fleshed', 'Lalit', 'Lalima', 'Red Supreme', 'Red-flHybr', 'Banarasi Surkha', 'Chittidar', 'Harijha', 'Sardar', 'Arka Mridula', 'Arka Amulya', 'Arka Kiran', 'Nagpur seedless', 'Hisar Surkha'	Dinesh and Vasugi 2010
Indonesia	'Kristal', 'Bangkok merah', 'Kamboja', 'Dadu1', 'Dadu 2', 'Pipit', 'Susupith'	Mayadewi et al. 2016
Malaysia	Jambubiji', 'Gu-5', 'Beaumont Semeyih', 'Beaumont Sungkai', 'Hongkong Pink'	Mitra and Thingreingam Irenaeus 2016
Mexico	'Media China', 'China'	Mitra and Thingreingam Irenaeus 2016
Pakistan	'Safeda', 'Allahabad', 'Lucknow-49', 'Red Fleshed', 'Seedless', 'Kerala', 'Apple Colour'	Mitra and Thingreingam Irenaeus 2016
South Africa	'Fan Retief', and 'TS-G2', 'Glom Toon' 'Klau', 'Khao Boon Soom Vietnam Xalynghe', 'Ruothong da lang'	Mitra and Thingreingam Irenaeus 2016; Rajan and Hudedamani 2019
Taiwan	'Jen-Ju', 'Di Wan', 'Pearl', 'Rainbow', 'Media China', 'China'	Mitra and Thingreingam Irenaeus 2016
Thailand	'Kim Ju', 'Pan Si Thong', 'Sa Li Thong'	Mitra and Thingreingam Irenaeus 2016
USA	'Homestead', 'Barbi Pink', 'Blitch', 'Hong Kong Pink', 'Patillo', 'Crystal', 'Lotus', 'Supreme', 'Webber', 'Beaumont', 'KaHua Kula'	Mitra and Thingreingam Irenaeus 2016
Vietnam	'Le Dai loan', 'Nu hoang', 'Khong hat', 'Xa li nghe', 'Ruot do', 'Se', 'Tim'	Mitra and Thingreingam Irenaeus 2016

of genes and with less environmental influence, which makes them a suitable candidate for improvement and positive selection (Pelea et al. 2012). Seed hardiness, which is an important fruit trait, is negatively correlated with the number of seeds per fruit but positively correlated with 100 seed weight of the fruit (Rajan et al. 2012).

Table 3 Some guava cultivars rich in nutritional values

HR traits	Cultivar name	Reference
High vitamin C	'Jen Ju', 'Di Wan' and 'Rainbow', 'Supreme', 'Jambubiji', 'Allahabad Safeda', 'Fan Retief', 'Lucknow-49', 'Media china', 'Sangam', 'Cotorrera', 'Microguayaba', 'H-118', 'H-138', 'H-153', ' H-345', 'Selection 106', 'Selection 126', 'BG 76-8', 'BG 76-14', 'Darío 18-2', 'BG 76-21', 'Red fleshed', 'Seedless', 'G. Vilas Pasand', 'Pant Prabhat', 'Mirzapur seedling', 'EC 162904', 'G-6', 'Chakaiya Ruthmanagar', 'Dhareedar'	Mitra and Thingreingam Irenaeus 2016; Medina and Herrero 2016; Dinesh and Vasugi 2010; Pommer 2012
Shelf life	'Yilan Red', 'Arka Amulya', 'Sardar', 'Swetha', 'Arka Mridula'	Mitra and Thingreingam Irenaeus 2016; Dinesh and Vasugi 2010
High Lycopene content	'Arka Kiran'	Mitra and Thingreingam Irenaeus 2016
High Phenolic content	'Allahabad Safeda ', 'Fan Retief', 'Lucknow-49', 'Sangam'	Medina and Herrero 2016
High Carotenoid content	'Cortibel 1'	Medina and Herrero 2016
High dietary fiber	'Allahabad Safeda ', 'Lucknow-49', 'Sangam', 'Lalit'	Medina and Herrero 2016
High Pectin content	'Allahabad Safeda ', 'Arka Mridula' 'Lucknow-49', 'Sangam', 'Lalit', Selected genotype no. 2'	Medina and Herrero 2016

It is interesting to see that fruit characters like fruit yield, number of fruits and fruit mass are unaffected by vegetative characters like plant height, stem diameter, and canopy volume. In contradiction, a high genetic correlation was observed between fruit yield with number of fruits; fruit weight with fruit length, width, fruit core diameter, number of seeds per fruit and seed weight which facilitate concurrent selection for these traits (Santos et al. 2017; Shiva 2017). Moreover, the total chlorophyll content has a positive correlation with Vitamin C content of the guava fruit. In case of leaf characters, leaf length and breadth had a positive correlation with leaf area, but were negatively correlated to stomata number (Shiva 2017). Further, high genotypic variance was observed in terms of width of leaf, whereas low variability for leaf length and length/width ratio (Nagar et al. 2018). More number of studies on heritability and genetic gain of important fruit traits with health benefits like vitamin C, TSS, phenolic and pectin content, flavonoids, and carotenoids are needed to improve the available germplasms and for the selection of progenies. Overall, the traditional breeding approaches in guava take a long period and require large parcels of land for progenies. Hence, adapting modern breeding

methods employing molecular markers and prediction tools will have an immense impact in terms of progeny screening, marker-assisted selection (MAS), genetic diversity analysis, association studies, genome editing, and genetic engineering.

5 Analysis of Genetic Diversity

Assessing the genetic diversity of guava germplasm can aid in crop improvement for traits of our interest, especially in the selection of parents for effective hybridization programs. Diversity analysis using phenotypic, biochemical, and genetic methods have been applied to study the available guava genetic resources in different countries (Singh et al. 2015b; Usman et al. 2020; Krause et al. 2021). Phenotypic characterization using descriptors related to tree, leaf, fruit, seed, and yield characters have assisted in understanding the divergence of guava germplasm and to identify suitable cultivars to enhance the genetic gain in breeding programs (Krause et al. 2021; Singh et al. 2015b; Pavani et al. 2022). Further, digital phenotyping has been applied to study the seed characters like texture, geometry, and color through digital image analysis to understand the genetic diversity among inbred guava lines (da Silva et al. 2021; Krause et al. 2017). Biochemical characterization consisting of health-related traits like Total soluble solids (TSS), total acidity, total sugars, sugar acid ratio, total flavonoids, total phenols, ascorbic acid, β-carotene, lycopene, and anti-oxidant activity have been studied to apprehend the divergence pattern among different white pulped cultivars (Usman et al. 2020) and different guava accessions (Santos et al. 2011).

Although many morphological characters are used for discriminating genotypes, they fail to differentiate between closely related genotypes. Thus, DNA markers or molecular markers were utilized for genetic diversity analysis of guava based on different origins, pulp color, vitamin C content, and cultivars. Markers like PCR-based, hybridization-based, and sequence-based have been used for diversity studies. Markers like Amplified Fragment Length Polymorphism (AFLP) (Thaipong et al. 2017), Random Amplified Polymorphic DNA (RAPD) (Ahmed et al. 2011), Inter simple sequence repeat (ISSR) (Sharaf et al. 2020), Inter-primer binding site (iPBS) (Mehmood et al. 2016), and sequence-related amplified polymorphism (SRAP) (Youssef and Ibrahim 2016) have been used to study the variability among the guava cultivars. Simple Sequence Repeat (SSR) markers have been predominantly used to study the genetic diversity of accessions/germplasms in different countries like India (Kumari et al. 2018; Kherwar et al. 2018), Kenya (Chiveu et al. 2019), China (Ma et al. 2020), United States (Sitther et al. 2014), Pakistan (Kareem et al. 2018), Bangladesh (Alam et al. 2018), Mexico (Sánchez-Teyer et al. 2010), and Venezuela (Aranguren et al. 2008). Genomic SSRs developed from cultivar 'Allahabad Safeda' through microsatellite-enriched libraries were found suitable for genetic diversity studies, population structure analysis and also had high cross-species transferability among *Psidium* species (Kumar et al. 2020). SSRs developed from *Psidium guajava* have had a high cross transferability potential to members of Myrtaceae family like *E. citriodora*, *E. camaldulensis*,

C. lanceolatus, and *S. aromaticum*, which can be utilized for inter- and intra-species genetic diversity analysis (Rai et al. 2013). Further, SSR markers have found its utility in studies assessing sexual compatibility among the *Psidium* species (Mulagund et al. 2021) and DNA fingerprinting (Chaithanya et al. 2017). Lately, due to advancements in next generation sequencing (NGS) and construction of draft genome assembly in guava, development of sequencing-based single nucleotide polymorphisms (SNPs), InDels, and transcriptome-based SSR markers from different peel and pulp-colored cultivars have become feasible. These markers have high applicability in population structure analysis, genetic diversity studies, and in breeding programs (Thakur et al. 2021). In addition, SNP markers identified using DArTSeq-based genotyping among different native Psidium species have shown high interspecific diversity (Grossi et al. 2021). The development of SNP-PCR and transcriptome-based SSR markers strongly associated with health-related traits can aid in precise identification of the cultivars, accessions, and wild species suitable for breeding programs in guava.

6 Molecular Mapping and QTL Identification for HR Genes

Various molecular markers like AFLP, SSR, RAPD, and SRAP were used for molecular mapping studies in guava. Genetic linkage maps were developed by mapping populations derived from heterozygous parents. Polymorphic markers were used to construct individual parental linkage maps and then assembled into an integrated map (Lepitre et al. 2010). Further, integration of phenotypic data with the marker data has helped in identifying QTLs that are correlated to different quantitative traits in guava (Ritter et al. 2010a).

The first molecular linkage map of guava contained 167 AFLP markers in the integrated map, which helped in the identification of 15 QTLs for vegetative characters (Valdés-Infante et al. 2003) (Table 4). A genetic linkage map utilizing 220 polymorphic AFLP markers was constructed along with the identification of QTLs for leaf length, leaf width, petiole length, height, fruit weight, acidity, total soluble solids, vitamin C content, and pulp thickness utilizing a mapping population derived from dwarf cultivar 'Enana Rojacubana' as the female parent (Rodriguez et al. 2005). Later, Ritter et al. (2010a) constructed an integrated parental linkage map in three guava mapping populations ('Enana' × 'N', 'Enana' × 'Suprema Roja' and 'Enana × 'Belic L-207') using AFLP and SSR markers. They obtained 11 linkage groups (LGs) corresponding to the haploid guava genome of 11 chromosomes for each parent. The integrated maps contained markers between 408 and 850 and were covering 1885–2179 cM, in length, respectively. Further, Ritter et al. (2010b) had conducted a QTL analysis using the linkage maps constructed for three mapping populations in their previous study. Sixteen different traits related to leaf and fruit characteristics and yields were recorded in both the parents and progenies of the three-mapping population. A total of 75, 56, and 59 QTLs for all the traits studied were detected in populations MP1 ('Enana' × N'), MP2 ('Enana' × 'Suprema Roja') and MP3 ('Enana' × 'Belic L-207'), respectively. Some QLTs were found

Table 4 Genetic linkage maps and QTLs identified in guava

Mapping population derived from crosses	Number of segregating individuals	Type of linkage map	Molecular markers employed	No. of QTLs identified	References
'Enana' × 'N6'	81	Integrated	AFLP	15 QTLs for vegetative traits	Valdés-Infante et al. (2003)
'Enana' × 'N6'	81	Integrated	SSR	Fruit weight – 3, fruit width – 3, acidity – 3, TSS – 2, Vit. C content – 2, pulp thickness – 2, seed number – 1, seed weight – 5	Rodriguez et al. (2005)
'Enana' × 'N6' 'Enana' × 'Suprema Roja' and 'Enana' × 'Belic L-207'	100–120	Integrated	AFLP and SSR	75, 56, and 59 QTLs in each mapping population respectively	Ritter et al. (2010a, b)
'Enana' × 'N6'	80	Integrated	AFLP and SSR	–	Lepitre et al. (2010)
'Kamsari' × 'Purple local'	94	Parental	SSR, SRAP, and RAPD	One major QTL for average fruit weight, four QTLs for seed strength (hardness/ softness)	Padmakar et al. (2015, 2016)

to be linked to SSR markers, which could be used for marker-assisted selection of progenies with desired traits in conventional breeding.

A similar type of study was conducted by Lepitre et al. (2010) in a guava mapping population derived from a cross between two heterozygous guava cultivars ('Enana' × 'N6'). A total of 1364 AFLP and SSR markers were used for linkage mapping. The integrated map length was 2179 cM and had an average linkage group length of 198 cM/chromosome. Further, high throughput genotyping of 94 guava F1 progenies using SSR and SRAP markers had generated scorable polymorphic markers, that were used for the construction of parental (Cultivar 'Kamsari' and 'Purple local') genetic linkage maps using a LOD score of 4.0 (Padmakar et al. 2015). Padmakar et al. (2016) have also identified one major QTL linked to fruit weight and two major and two minor QTLs for seed strength based on linkage map enriched further using RAPD markers. These studies have shown some insight into the QTL regions responsible for health related traits like vitamin C, TSS, and acidity, along with other fruit and leaf-related quantitative and qualitative traits. However,

adapting NGS-based genotyping methods like Genotyping by sequencing (GBS) becomes vital for exact identification of QTLs or genes linked to health-related traits.

In addition, an RNA-Seq based transcriptome analysis of genotypes with contrasting peel and pulp colors at different stages of ripening has set the stage for the functional genomics of traits like colored pulp and peel (Mittal et al. 2020) in guava. The study has identified 68 candidate genes that are involved in peel and pulp color development in guava fruits. These genes involved in ethylene biosynthesis, phenylpropanoid, and monolignol pathways, along with the QTLs identified, have the potential to be used as predictive markers in marker-assisted selection (MAS) in conventional breeding. Interestingly, even though there is a dearth of NGS-based research in this crop, guava got its first chromosome-level genome assembly for the cultivar 'New Age' recently (Feng et al. 2021). This development will aid in reference-based genotyping, thereby assisting in construction of high-density linkage map, genome-wide association studies (GWAS), and Genomic selection (GS) in guava.

7 Nanotechnology

Nanoparticles (NPs) are produced through a complex physico-chemical reaction which involves release of toxic wastes, unsafe by-products, and a high level of energy intake. Hence, an alternate and safe method for metal nanoparticle synthesis can be achieved through green biosynthesis of NPs utilizing plant sources. Guava leaves rich in phytochemicals like flavonoids, phenols, polyphenols, tannins, saponins, terpenoids, ascorbic acid, and proteins can mediate the redox reactions, thereby reducing the salts/ions, stabilizing and capping the nanoparticles formed. Several metal nanoparticles synthesized using guava leaf extracts were reported to possess antimicrobial properties. The aqueous extracts of guava leaves have been used in the preparation of iron oxide, silver oxide (Dildar et al. 2022), and Zinc oxide (Jyoti et al. 2020; Balalakshitha and Kolanjinathan 2021; Ramya et al. 2022) nanoparticles that exhibited radical scavenging and antimicrobial activities against pathogenic Gram positive and Gram negative bacterial strains. Similarly, the silver nanoparticles synthesized from guava extract coated with polymeric micelles possessed high antifungal activity against *Candida albicans* (Suwan et al. 2019). In addition, these silver NPs presented high level of antioxidant, antibacterial, and cytotoxicity against the colon cancer cells (Sandhiya et al. 2021). The copper NPs synthesized using aqueous guava extract comprised of high antibacterial activity against bacterial pathogens, *E. coli* and *S. aureus* due to the presence of antioxidant compounds like polyphenols and ascorbic acid in the guava extract (Caroling et al. 2015). Likewise, aqueous and ethanol extracts of guava leaves have yielded gold NPs due to the involvement of flavonoids present in the leaf extracts (Raghunandan et al. 2009; Putri et al. 2021). Furthermore, a lipophilic nano-emulsion containing purified lycopene from red guava had shown potential as a drug delivery system because

of its successful delivery of the lycopene to liver, kidney, and prostate organs of mice and also for enhancing the cytotoxicity against the prostate cancer cells (Vasconcelos et al. 2021).

Another interesting application of nanotechnology is for the improvement of shelf life of guava fruits using nanoparticle-based coating materials. The edible coatings, like Zinc oxide NPs combined with chitosan and alginate (Arroyo et al. 2020) and solid lipid NPs (Zambrano-Zaragoza et al. 2013), were shown to increase the shelf life of the guava fruits. In addition, incorporation of guava puree in chitosan nanoparticle-based edible films for packaging has imparted favorable color and odor for these films (Lorevice et al. 2012). The nano-technological application of guava leaf extracts in synthesis of metal NPs can prove to be a cost-effective and safe method with a wide biomedical application.

8 Social, Political, and Regulatory Issues

The health benefits of guava plants can immensely contribute to the nutritional security of the growing population. Unfavorable weather conditions, pests and diseases, unavailability of good commercial cultivars, lack of processing and storage facilities, and post-harvest wastages are the major problems faced by guava growers. Awareness toward the market potential of guava fruit is needed to boost its cultivation among the farmers of the tropical and subtropical countries. Moreover, popularizing guava fruit and its benefits among the public can provide consumers with a cheaper and nutritious alternative. More importantly, developing post-harvest techniques and value-added products that retain the phytochemicals and antioxidants present in the fresh fruit could be a boon to the industry and consumer market. In order to ensure its safe deployment, regulatory issues related to its biomedical application, nanotechnology, and alternative therapies need to be streamlined.

9 Future Prospects

The tropical and subtropical regions are home to *Psidium guajava*, which is highly diverse and spread across the world. Studies have proved the antioxidant, anticancerous, antidiabetic, antidiarrheal, and anti-inflammatory potential of this highly nutritious commercial fruit crop. Although there are a few conventional breeding efforts, application of modern breeding techniques using precise molecular markers for marker-assisted selection (MAS), marker-assisted introgression (MAI), association studies, genomic selection, genome editing, and transformation studies are lacking in guava. Using high-throughput sequencing techniques like NGS can help in identifying SNP and SSR markers for important health-related traits in guava and thereby assist in speed breeding of superior guava cultivars and hybrids. Also, there is a need to develop strong international collaborations, industrial and academic interactions to ensure adequate public and private funding of projects which helps in faster outcome.

10 Conclusion

Nutraceutical properties present in guava plant is being utilized in the traditional medicine practices of different countries. From leaf to seed, the potential benefits provided by this underutilized crop can help cure and mitigate various ailments as documented by different studies. On the other hand, it is essential to develop breeding materials and value-added products and popularize its benefits among the general public for a sustainable growth in guava production and utilization.

References

Abdullah MS, Nas FS, Ali M (2019) Antibacterial activity of *Psidium guajava* leaf and stem bark extracts against clinical isolates of *Staphylococcus aureus* and *Salmonella typhi*. Int J Res Pharm Biosci 6(5):11–17

Ahmed B, Mannan MA, Hossain SA (2011) Molecular characterization of guava (*Psidium guajava* L.) germplasm by RAPD analysis. Int J Nat Sci 1(3):62–67

Alam F, Islam KD, Rahman SM (2018) Variability among selective guava (*Psidium guajava* L.) varieties revealed by morphology and RAPD marker. Jahangirnagar Univ J Biol Sci 7(2):89–98

Alhamdi H (2019) Anticancer potential of guava leaves extract against breast, lung and prostate cancer cell lines. Doctoral dissertation, Tennessee State University

Alvarez-Suarez JM, Giampieri F, Gasparrini M, Mazzoni L, Forbes-Hernández TY, Afrin S, Battino M (2018) Guava (*Psidium guajava* L. cv. Red Suprema) crude extract protect human dermal fibroblasts against cytotoxic damage mediated by oxidative stress. Plant Foods Hum Nutr 73(1): 18–24

Anjali K, Manjul SP (2021) Study on formulation and evaluation of famotidine tablets prepared by using guava starch as binding agents. Curr Asp Pharm Res Dev 5:59–66

Aranguren Y, Briceño A, Fermin G (2008) Assessment of the variability of Venezuelan guava landraces by microsatellites. In: II international symposium on guava and other Myrtaceae, vol 849, pp 147–154

Arroyo BJ, Bezerra AC, Oliveira LL, Arroyo SJ, de Melo EA, Santos AMP (2020) Antimicrobial active edible coating of alginate and chitosan add ZnO nanoparticles applied in guavas (*Psidium guajava* L.). Food Chem 309:125566

Balalakshitha M, Kolanjinathan K (2021) Phytochemistry and bioactivity of green synthesized nanoparticle from *Psidium guajava* against selected diabetic wound pathogens. Bull Environ Pharmacol Life Sci 10:115–119

Birdi T, Krishnan GG, Kataria S, Gholkar M, Daswani P (2020) A randomized open label efficacy clinical trial of oral guava leaf decoction in patients with acute infectious diarrhoea. J Ayurveda Integr Med 11(2):163–172

Biswas B, Rogers K, McLaughlin F, Daniels D, Yadav A (2013) Antimicrobial activities of leaf extracts of guava (*Psidium guajava* L.) on two gram-negative and gram-positive bacteria. Int J Microbiol. https://doi.org/10.1155/2013/746165

Blancas-Benitez FJ, Pérez-Jiménez J, Sañudo-Barajas JA, Rocha-Guzmán NE, González-Aguilar GA, Tovar J, Sáyago-Ayerdi SG (2022) Indigestible fraction of guava fruit: phenolic profile, colonic fermentation and effect on HT-29 cells. Food Biosci 2022:101566

Bogha TT, Sawate AR, Kshirsagar RB, Bochare SS (2020) Studies on physical, chemical and mineral evaluation of guava (*Psidium Guajava* L.). Pharm Innov J 9(3):117–119

Caroling G, Priyadharshini MN, Vinodhini E, Ranjitham AM, Shanthi P (2015) Biosynthesis of copper nanoparticles using aqueous guava extract-characterisation and study of antibacterial effects. Int J Pharm Biol Sci 5(2):25–43

Chaithanya N, Sailaja D, Dinesh MR, Vasugi C, Lakshmana Reddy DC, Aswath C (2017) Microsatellite-based DNA fingerprinting of guava (*Psidium guajava*) genotypes. Proc Natl Acad Sci India Sect B Biol Sci 87(3):859–867

Chang YP, Woo KK, Gnanaraj C (2020) Pink guava. In: Valorization of fruit processing by-products. Academic Press, London, pp 227–252

Chauhan AK, Singh S, Singh RP, Singh SP (2015) Guava-enriched dairy products: a review. Indian J Dairy Sci 68(1):01–05

Chiveu JC, Mueller M, Krutovsky KV, Kehlenbeck K, Pawelzik E, Naumann M (2019) Genetic diversity of common guava in Kenya: an underutilized naturalized fruit species. Fruits 74(5): 236–248

da Silva Lima R, Ferreira SRS, Vitali L, Block JM (2019) May the superfruit red guava and its processing waste be a potential ingredient in functional foods? Food Res Int 115:451–459

da Silva CCA, Vieira HD, Viana AP, Santos EA, Maitan MQ (2021) Digital phenotyping in inbred guava lines: seed characterization. Funct Plant Breed J 3(2):33–50

Denny C, Melo PS, Franchin M, Massarioli AP, Bergamaschi KB, de Alencar SM, Rosalen PL (2013) Guava pomace: a new source of anti-inflammatory and analgesic bioactives. BMC Complement Altern Med 13(1):1–7

Dildar N, Ali SN, Sohail T, Lateef M, Khan ST, Bukhari SF, Fazil P (2022) Biosynthesis, characterization, radical scavenging and antimicrobial properties of *Psidium guajava* Linn coated silver and iron oxide nanoparticles. Egypt J Chem 65(2):145–151

Dinesh MR, Vasugi C (2010) Guava improvement in India and future needs. J Hortic Sci 5(2):94–108

Dinesh M, Bharathi K, Vasugi C, Veena GL, Ravishankar KV, Nischita P (2017) Inheritance studies and validation of hybridity in guava (*Psidium guajava*). Indian J Agric Sci 87:42–45

Dos Santos RC, Ombredane AS, Souza JMT, Vasconcelos AG, Plácido A et al (2018) Lycopene-rich extract from red guava (*Psidium guajava* L.) displays cytotoxic effect against human breast adenocarcinoma cell line MCF-7 via an apoptotic-like pathway. Food Res Int 105:184–196

Elabd M (2018) Effect of antimicrobial edible coatings on quality and shelf life of minimal processed guava (*Psidium guajava*). Alex J Food Sci Technol 15(1):65–76

Emanuel N, Sao K, Kaushik A (2018) Bioactive compounds, antioxidant properties, and metal content studies of guava fruit by-products for value added processing. Braz J Anal Chem 5(21):8–18

Etemadipoor R, Dastjerdi AM, Ramezanian A, Ehteshami S (2020) Ameliorative effect of gum arabic, oleic acid and/or cinnamon essential oil on chilling injury and quality loss of guava fruit. Sci Hortic 266:109255

Feng C, Feng C, Lin X, Liu S, Li Y, Kang M (2021) A chromosome-level genome assembly provides insights into ascorbic acid accumulation and fruit softening in guava (*Psidium guajava*). Plant Biotechnol J 19(4):717–730

Grossi LL, Fernandes M, Silva MA, de Oliveira Bernardes C, Tuler AC et al (2021) DArTseq-derived SNPs for the genus *Psidium* reveal the high diversity of native species. Tree Genet Genomes 17(2):1–13

Guintu FZ, Chua AH (2013) Effectivity of guava leaves (*Psidium guajava*) as mouthwash for patients with aphthous ulcers. Philip J Otolaryngol Head Neck Surg 28(2):8–13

Gupta N, Kour A (2019) Genetic parameters, character association and path analysis for fruit yield and its component characters in guava (*Psidium guajava* L.). Electron J Plant Breed 10(1):256–263

Gurmachhan CM, Tandukar U, Shrestha N, Bahadur P (2020) Antimicrobial and phytochemical studies of methanolic bark extract of *Psidium guajava* L. and *Punica granatum* L. J Plant Resour 18(1):211

Hoyos CG, Guerra AS, Pérez SA, Velásquez-Cock J, Villegas M, Gañán P, Gallego RZ (2022) An edible oil enriched with lycopene from pink guava (*Psidium guajava* L.) using different mechanical treatments. Molecules 27(3):1038

Iha OK, Martins GB, Ehlert E, Montenegro MA, Sucupira RR, Suarez PA (2018) Extraction and characterization of passion fruit and guava oils from industrial residual seeds and their application as biofuels. J Braz Chem Soc 29:2089–2095

Irondi EA (2020) Phenolics-rich extract of guava stem bark inhibits enzymes associated with nephrolithiasis and obesity in vitro. J Complement Med Res 11(1):34–42

Jaleel W, Li Q, Shi Q, Qi G, Latif M, Ali S, Yasin NA, Lyu L, He Y (2021) Using GCMS to find out the volatile components in the aroma of three different commercial fruits in China. J Anim Plant Sci 31(1):12

Jamesa JI, Aisami A, Muhammad SA, Saidu L, Patricia S, Peter ML (2020) *Psidium guajava* seed protein hydrolysates exhibit in vitro antioxidant and inhibitory activity against α-amylase and α-glucosidase. Int J Innov Approach Sci Res 4(4):84–99

Jayachandran M, Vinayagam R, Ambati RR, Xu B, Chung SSM (2018) Guava leaf extract diminishes hyperglycemia and oxidative stress, prevents β-cell death, inhibits inflammation, and regulates NF-kB signaling pathway in STZ induced diabetic rats. Biomed Res Int. https://doi.org/10.1155/2018/4601649

Jiang LR, Qin Y, Nong JL, An H (2021) Network pharmacology analysis of pharmacological mechanisms underlying the anti-type 2 diabetes mellitus effect of guava leaf. Arab J Chem 14 (6):103143

Jyoti A, Agarwal M, Tomar RS (2020) Synergistic effect of zinc oxide nanoparticles and guava leaf extract for enhanced antimicrobial activity against enterotoxigenic *Escherichia coli*. J Biochem Technol 11(2):17–23

Kalia A, Kaur M, Shami A, Jawandha SK, Alghuthaymi MA et al (2021) Nettle-leaf extract derived ZnO/CuO nanoparticle-biopolymer-based antioxidant and antimicrobial nanocomposite packaging films and their impact on extending the post-harvest shelf life of guava fruit. Biomolecules 11(2):224

Kamble M, Selwate T, Dhabarde D, Ingole A, Baheti J (2019) Formulation and evaluation of anti-acne face wash gel using guava seed extract. J Drug Deliv Therap 9(3):5–7

Kareem A, Jaskani MJ, Mehmood A, Khan IA, Awan FS, Sajid MW (2018) Morpho-genetic profiling and phylogenetic relationship of guava (*Psidium guajava* L.) as genetic resources in Pakistan. Rev Bras Frutic 40:120

Kaushik A (2019) Effects of developed guava seeds biscuit on blood glucose level of type 2 diabetes patients. In: International conference on food security through agriculture & allied sciences, vol 5, pp 57–58

Kherwar D, Usha K, Mithra SV, Singh B (2018) Microsatellite (SSR) marker assisted assessment of population structure and genetic diversity for morpho-physiological traits in guava (*Psidium guajava* L.). J Plant Biochem Biotechnol 27(3):284–292

Kim TY (2011) Antiproliferation and dedifferentiation in thyroid cancer cell lines by polyphenol phytochemicals. J Korean Med Sci 26:893–899

König A, Schwarzinger B, Stadlbauer V, Lanzerstorfer P, Iken M et al (2019) Guava (*Psidium guajava*) fruit extract prepared by supercritical CO_2 extraction inhibits intestinal glucose resorption in a double-blind, randomized clinical study. Nutrients 11(7):1512

Krause W, Viana AP, Cavalcante NR, Ambrósio M, Santos EA, Vieira HD (2017) Digital phenotyping for quantification of genetic diversity in inbred guava (*Psidium guajava*) families. Genet Mol Res 16(1):gmr16019474. https://doi.org/10.4238/gmr16019474

Krause W, Cavalcante NR, Viana AP, Ambrósio M, Santos EA, Vieira HD (2021) Genetic variability in inbred guava families mediated by agronomic traits. Funct Plant Breed J 3(2):7

Kuber BR, Lakshmi MR, Deepika E, Yamini P (2013) Phytochemical screening, in vitro antibacterial and antioxidant activity of the *Psidium guajava* root bark. Int J Curr Microbiol Appl Sci 2(10):238–248

Kumar V, Gupta P (2021) Studies on preparation of guava ice cream from (CV.) Allahabad Safeda and L-49 guava (*Psidium guajava* L.). Progress Agric 21(1):78–81

Kumar C, Kumar R, Singh SK, Goswami AK, Nagaraja A, Paliwal R, Singh R (2020) Development of novel g-SSR markers in guava (*Psidium guajava* L.) cv. Allahabad Safeda and their application in genetic diversity, population structure and cross species transferability studies. PLoS One 15(8):e0237538

Kumar M, Kapoor S, Dhumal S, Tkaczewska J, Changan S et al (2022) Guava (*Psidium guajava* L.) seed: a low-volume, high-value byproduct for human health and the food industry. Food Chem 2022:132694

Kumari S, Nagaraja Arumugam RS, Srivastav M, Banoth S et al (2018) Diversity analysis of guava (*Psidium guajava*) germplasm collection. Indian J Agric Sci 88:489–497

Laily N, Kusumaningtyas RW, Sukarti I, Rini MRDK (2015) The potency of guava *Psidium guajava* (L.) leaves as a functional immunostimulatory ingredient. Procedia Chem 14:301–307

Lepitre V, Nansot G, Grangeon R, Pomies V, Rivallan R et al (2010) The microsatellite (SSR)/AFLP reference linkage map of guava. Acta Hortic 849:183–192

Lin HC, Lin JY (2020) GSF3, a polysaccharide from guava (*Psidium guajava* L.) seeds, inhibits MCF-7 breast cancer cell growth via increasing Bax/Bcl-2 ratio or Fas mRNA expression levels. Int J Biol Macromol 161:1261–1271

Lin HC, Lin JY (2021) Pharmacological effects of guava (*Psidium guajava* L.) seed polysaccharides: GSF3 inhibits PC-3 prostate cancer cell growth through immunotherapy in vitro. Int J Mol Sci 22(7):3631

Liu HC, Chiang CC, Lin CH, Chen CS, Wei CW, Lin SY, Yiang GT, Yu YL (2020) Anti-cancer therapeutic benefit of red guava extracts as a potential therapy in combination with doxorubicin or targeted therapy for triple-negative breast cancer cells. Int J Med Sci 17(8):1015

Lorevice MV, Moura MRD, Aouada FA, Mattoso LH (2012) Development of novel guava puree films containing chitosan nanoparticles. J Nanosci Nanotechnol 12(3):2711–2717

Luo Y, Peng B, Wei W, Tian X, Wu Z (2019) Antioxidant and anti-diabetic activities of polysaccharides from guava leaves. Molecules 24(7):1343

Ma Z, Liu S, Liang Z, Xu S, Hu W (2020) Analysis of genetic diversity of 45 guava germplasm evaluated using SSR markers. Int J Fruit Sci 20(3):385–393

Malo SE, Campbell CW (1994) The guava. IFAS extension. Bulletin, University of Florida

Mangaraj S, Thakur RR, Mathangi RS, Yadav A, Swain S (2021) Shelf life enhancement of guava (*Psidium guajava* cv. Baruipur) stored under pilot scale modified atmosphere storage system. Food Sci Technol Int 27(7):674–689

Mayadewi NNA, Rai IN, Dwiyani R, Astarini IA (2016) Genetic diversity of guava (*Psidium guajava* L.) in Bali, Indonesia based on RAPD marker. Int J Biosci Biotechnol 4(1):20

Medina NNR, Herrero JVI (2016) Guava (*Psidium guajava* L.) cultivars: an important source of nutrients for human health. In: Simmonds M, Preedy VR (eds) Nutritional composition of fruit cultivars. Academic Press, London, pp 287–315

Mehmood A, Luo S, Ahmad NM, Dong C, Mahmood T, Sajjad Y, Jaskani MJ, Sharp P (2016) Molecular variability and phylogenetic relationships of guava (*Psidium guajava* L.) cultivars using inter-primer binding site (iPBS) and microsatellite (SSR) markers. Genet Resour Crop Evol 63(8):1345–1361

Meles DK, Mustofa I, Wurlina W, Susilowati S, Utama S, Suwasanti N, Putri DKSC (2021) The restorative effect of red guava (*Psidium guajava* L.) fruit extract on pulmonary tissue of rats (*Rattus norvegicus*) exposed to cigarette smoke. Vet Med Int 2021:9931001. https://doi.org/10.1155/2021/9931001

Melo C, Cornejal N, Cruz V, Alsaidi S, Cruz Rodriguez G et al (2020) Antioxidant capacity and antimicrobial activity of commercial samples of guava leaves (*Psidium guajava*). J Med Active Plants 9(1):2

Mitra SK, Thingreingam Irenaeus KS (2016) Guava cultivars of the world. In: International Symposia on Tropical and Temperate Horticulture – ISTTH 2016, vol 1205, pp 905–910

Mittal A, Yadav IS, Arora NK, Boora RS, Mittal M, Kaur P, Erskine W, Chhuneja P, Gill MIS, Singh K (2020) RNA-sequencing based gene expression landscape of *guava* cv. Allahabad Safeda and comparative analysis to colored cultivars. BMC Genomics 21(1):1–19

Mulagund J, Dinesh MR, Vasugi C, Rekha A, Ravishankar KV (2021) Studies on sexual compatibility and or incompatibility among *Psidium* species and their hybridity confirmation through SSR markers. Isr J Plant Sci 68(4):253–266

Murmu SB, Mishra HN (2018) The effect of edible coating based on Arabic gum, sodium caseinate and essential oil of cinnamon and lemon grass on guava. Food Chem 245:820–828
Musaa KH, Abdullaha A, Subramaniamb V (2015) Flavonoid profile and antioxidant activity of pink guava. Skin 172(36b):308–57a
Nagar PK, Satodiya BN, Prajapati DG, Nagar SK, Patel KV (2018) Assessment of genetic variability and morphological screening of guava (*Psidium guajava* L.) hybrids. Pharm Innov J 7(8):35–40
Nwaichi EO, Chuku LC, Oyibo NJ (2015) Profile of ascorbic acid, beta-carotene and lycopene in guava, tomatoes, honey and red wine. Int J Curr Microbiol Appl Sci 4(2):39–43
Oncho DA, Ejigu MC, Urgessa OE (2021) Phytochemical constituent and antimicrobial properties of guava extracts of east Hararghe of Oromia, Ethiopia. Clin Phytosci 7(1):1–10
Padmakar B, Kanupriya C, Latha PM, Prashant KS, Dinesh MR, Sailaja D, Aswath C (2015) Development of SRAP and SSR marker-based genetic linkage maps of guava (*Psidium guajava* L.). Sci Hortic 192:158–165
Padmakar B, Kanupriya C, Latha PM, Vasugi C, Dinesh MR, Sailaja D, Aswath C (2016) Enrichment of genetic linkage maps and mapping QTLs specific to seed strength-hardness/softness-in guava (*Psidium guajava* L.). J Hortic Sci 11(1):13–20
Pandian RS, Jayalakshmi M (2019) HPLC analysis of water soluble vitamin B in leaves *Psidium guava*. Asian J Pharm Pharmacol 5(1):69–72
Patel RK, Maiti CS, Deka BC, Vermav VK, Deshmukh NA, Verma MR (2015) Genetic variability, character association and path coefficient study in guava (*Psidium guajava* L.) for plant growth, floral and yield attributes. Int J Bioresour Stress Manag 6(4):457–466
Pavani S, Tanuja Priya B, Madhumathi C (2022) Genetic divergence in guava (*Psidium guajava* L.) genotypes. Pharm Innov J 11(2):599–602
Pelea LP, González AS, Fernández EB, Rodríguez Medina NN, Valdés-Infante Herrero J, Pommer CV (2012) Heritability estimates of guava (*Psidium guajava* L.) agricultural important characters evaluated in three populations. In: III international symposium on guava and other Myrtaceae, pp 117–123
Pérez-Beltrán YE, Becerra-Verdín EM, Sáyago-Ayerdi SG, Rocha-Guzmán NE, García-López EG et al (2017) Nutritional characteristics and bioactive compound content of guava purees and their effect on biochemical markers of hyperglycemic and hypercholesterolemic rats. J Funct Foods 35:447–457
Phonarknguen R, Nobsathian S, Assawasuparerk K (2022) Effect of Betulinic acid extraction from guava (*Psidium guajava* Linn.) leaves against human cholangiocarcinoma cells. Asian Pac J Cancer Prev 23(2):583–590
Polinati RM, Teodoro AJ, Correa MG, Casanova FA, Passos CLA, Silva JL, Fialho E (2022) Effects of lycopene from guava (*Psidium guajava* L.) derived products on breast cancer cells. Nat Prod Res 36(5):1405–1408
Pommer CV (2012) Guava world-wide breeding: major techniques and cultivars and future challenges. In: III international symposium on guava and other Myrtaceae, vol 959, pp 81–88
Prommaban A, Utama-ang N, Chaikitwattana A, Uthaipibull C, Srichairatanakool S (2019) Linoleic acid-rich guava seed oil: safety and bioactivity. Phytother Res 33(10):2749–2764
Prommaban A, Utama-Ang N, Chaikitwattana A, Uthaipibull C, Porter JB, Srichairatanakool S (2020) Phytosterol, lipid and phenolic composition, and biological activities of guava seed oil. Molecules 25(11):2474
Putri SE, Pratiwi DE, Side S (2021) The effect of microwave irradiation on synthesis of gold nanoparticles using ethanol extract of white bol guava leaves. J Phys Conf Ser 1752(1):012058
Raghunandan D, Basavaraja S, Mahesh B, Balaji S, Manjunath SY, Venkataraman A (2009) Biosynthesis of stable polyshaped gold nanoparticles from microwave-exposed aqueous extracellular anti-malignant guava (*Psidium guajava*) leaf extract. Nanobiotechnology 5(1):34–41
Rai MK, Phulwaria M, Shekhawat NS (2013) Transferability of simple sequence repeat (SSR) markers developed in guava (*Psidium guajava* L.) to four Myrtaceae species. Mol Biol Rep 40 (8):5067–5071

Rajan S, Hudedamani U (2019) Genetic resources of guava: importance, uses and prospects. In: Rajasekharan PE, Ramanatha Rao V (eds) Conservation and utilization of horticultural genetic resources. Springer, Singapore, pp 363–383

Rajan S, Yadava LP, Kumar R (2012) Variation among guava (*Psidium guajava* L.) accessions in seed hardness and its association with fruit characteristics. Int J Innov Hortic 1(1):18–23

Rakmai J, Cheirsilp B, Mejuto JC, Simal-Gándara J, Torrado-Agrasar A (2018) Antioxidant and antimicrobial properties of encapsulated guava leaf oil in hydroxypropyl-beta-cyclodextrin. Ind Crop Prod 111:219–225

Ramya V, Kalaiselvi V, Kannan SK, Shkir M, Ghramh HA et al (2022) Facile synthesis and characterization of zinc oxide nanoparticles using *Psidium guajava* leaf extract and their antibacterial applications. Arab J Sci Eng 47(1):909–918

Ritter E, Herran A, Valdés-Infante J, Rodríguez-Medina MN, Briceño A et al (2010a) Comparative linkage mapping in three guava mapping populations and construction of an integrated reference map in guava. In: Rohde W, Fermin G (eds) II international symposium on guava and other Myrtaceae, vol 849, pp 175–182

Ritter E, Rodríguez-Medina NN, Velásquez B, Rivero D, Rodríguez JA et al (2010b) QTL (quantitative trait loci) analysis in guava. In: Rohde W, Fermin G (eds) II international symposium on guava and other Myrtaceae, vol 849, pp 193–202

Rodriguez N, Valdes-Infante J, Becker D, Velazquez B, Gonzalez G et al (2005) Characterization of guava accessions by SSR markers, extension of the molecular linkage map, and mapping of QTLs for vegetative and reproductive characters. In: Singh G et al. (eds) I international guava symposium, vol 735, pp 201–216

Rojas-Garbanzo C, Rodríguez L, Pérez AM, Mayorga-Gross AL, Vásquez-Chaves V et al (2021) Anti-platelet activity and chemical characterization by UPLC-DAD-ESI-QTOF-MS of the main polyphenols in extracts from *Psidium* leaves and fruits. Food Res Int 141:110070

Ruangsuriya J, Wongpoomchai R, Srichairatanakool S, Sirikul W, Buawangpong N, Siviroj P (2022) Guava fruit and *Acacia pennata* vegetable intake association with frailty of older adults in northern Thailand. Nutrients 14(6):1192

Sampath Kumar NS, Sarbon NM, Rana SS, Chintagunta AD, Prathibha S et al (2021) Extraction of bioactive compounds from *Psidium guajava* leaves and its utilization in preparation of jellies. AMB Express 11(1):1–9

Sánchez-Teyer LF, Barraza-Morales A, Quiroz-Moreno A, Ortiz-Garcia MM, Becerril-Chi K et al (2010) Genetic diversity of Mexican guava germplasm evaluated using AFLP and SSR markers. Acta Hortic 859:255–260

Sandhiya V, Gomathy B, Sivasankaran MR, Thirunavukkarasu P, Mugip Rahaman A, Asha S (2021) Green synthesis of silver nanoparticles from guava (*Psidium guajava* Linn.) leaf for antibacterial, antioxidant and cytotoxic activity on HT-29 cells (Colon cancer). Ann Roman Soc Cell Biol 25(6):20148–20163

Santos CAF, Corrêa LC, Costa SRD (2011) Genetic divergence among *Psidium* accessions based on biochemical and agronomic variables. Crop Breed Appl Biotechnol 11:149–156

Santos PRD, Preisigke SDC, Viana AP, Cavalcante NR, Sousa CMBD, Amaral ATD (2017) Associations between vegetative and production traits in guava tree full-sib progenies. Pesqui Agropecu Bras 52:303–310

Shabbir H, Kausar T, Noreen S, Rehman HU, Hussain A, Huang Q, Gani A, Su S, Nawaz A (2020) In vivo screening and antidiabetic potential of polyphenol extracts from guava pulp, seeds and leaves. Animals 10(9):1714

Shanthirasekaram K, Bulugahapitiya VP, Manawadu H, Gangabadage CS (2021) Phytochemicals and antioxidant properties of the leaves of wild guava varieties grown in Sri Lanka. J Sci 12(2): 33–46

Sharaf MM, Khamis MA, Bakry KHA, Saaed WT, EL-Tarawy OM (2020) Genetic identification of some guava seedling trees. Ann Agric Sci Moshtohor 58(4):1039–1050

Shiva B (2017) Correlation studies among vegetative, fruit physico–chemical characters of guava (*Psidium guajava* L.). Int J Agric Sci 9:4322. 0975-3710

Shukla S, Kushwaha R, Singh M, Saroj R, Puranik V, Agarwal R, Kaur D (2021) Quantification of bioactive compounds in guava at different ripening stages. Food Res 5(3):183–189

Singh D, Gill MIS, Boora RS, Arora NK (2015a) Estimates of genetic variability, heritability, genetic advance, correlation coefficients and their prospects for crop improvement in guava (*Psidium guajava* L.). J Appl Hortic 17(1):76–78

Singh D, Gill MIS, Boora RS, Arora NK (2015b) Genetic diversity analysis in guava (*Psidium guajava*) on the basis of morphological and physico-chemical traits. Indian J Agric Sci 85(5): 678–683

Sitther V, Zhang D, Harris DL, Yadav AK, Zee FT, Meinhardt LW, Dhekney SA (2014) Genetic characterization of guava (*Psidium guajava* L.) germplasm in the United States using microsatellite markers. Genet Resour Crop Evol 61(4):829–839

Soliman FM, Fathy MM, Salama MM, Saber FR (2016) Comparative study of the volatile oil content and antimicrobial activity of *Psidium guajava* L. and *Psidium cattleianum* Sabine leaves. Bull Fac Pharm Cairo Univ 54(2):219–225

Sormin RE, Ria MB, Ola CYI (2020) Relative effectiveness of red guava fruit (*Psidium guajava* L.) and star fruit (*Averrhoa carambola* L.) on hemoglobin level in pregnant women at Camplong Community Health Center, Kupang District. In: 7th international conference on public health, Sebelas Maret University, pp 49–55

Sriwilaijaroen N, Fukumoto S, Kumagai K, Hiramatsu H, Odagiri T, Tashiro M, Suzuki Y (2012) Antiviral effects of *Psidium guajava* Linn. (guava) tea on the growth of clinical isolated H1N1 viruses: its role in viral hemagglutination and neuraminidase inhibition. Antiviral Res 94(2): 139–146

Subramanyam MD, Iyer CPA (1990) Studies on inheritance in guava (*Psidium guajava* L.). Fruit Breed Genet 317:255–258

Suwan T, Khongkhunthian S, Okonogi S (2019) Antifungal activity of polymeric micelles of silver nanoparticles prepared from *Psidium guajava* aqueous extract. Drug Discov Ther 13(2): 62–69

Suwanwong Y, Boonpangrak S (2021) Phytochemical contents, antioxidant activity, and anticancer activity of three common guava cultivars in Thailand. Eur J Integr Med 42:101290

Tan S, Wang Z, Xiang Y, Deng T, Zhao X, Shi S, Zheng Q, Gao X, Li W (2020) The effects of drying methods on chemical profiles and antioxidant activities of two cultivars of *Psidium guajava* fruits. Food Sci Technol 118:108723

Thaipong K, Promchot S, Auvuchanon A, Boonprakob U (2017) Genetic analysis of guava germplasm using AFLP markers. Int J Agric Technol 13(5):741–752

Thakur S, Yadav IS, Jindal M, Sharma PK, Dhillon GS et al (2021) Development of genome-wide functional markers using draft genome assembly of guava (*Psidium guajava* L.) cv. Allahabad Safeda to expedite molecular breeding. Front Plant Sci 12:21

U.S. Department of Agriculture (USDA) (2019) Agricultural Research Service. Food Data Central. fdc.nal.usda.gov

Usman M, Zaman Q, Fatima B, Rana IA, Awan FS (2020) Morpho-chemical diversity and RAPD fingerprinting in white flesh guava cultivars. J Anim Plant Sci 30:398–409

Valdés-Infante D, Becker D, Rodríguez N, Velázquez B, González G, Sourd D, Rodríguez J, Ritter E, Rohde W (2003) Molecular characterization of Cuban accessions of guava (*Psidium guajava* L.) establishment of a first molecular linkage map and mapping of QTLs for vegetative characters. Indian J Genet Plant Breed 57:349–358

Vasconcelos AG, Barros ALA, Cabral WF, Moreira DC, da Silva IGM et al (2021) Promising self-emulsifying drug delivery system loaded with lycopene from red guava (*Psidium guajava* L.): in vivo toxicity, biodistribution and cytotoxicity on DU-145 prostate cancer cells. Cancer Nanotechnol 12(1):1–29

Xu L, Zhang L, Li YH, Li LY, Xie ZH, Tu ZC (2022) Inhibitory activity and mechanism of guavinoside B from guava fruits against α-glucosidase: insights by spectroscopy and molecular docking analyses. J Food Biochem 2022:e14101

Youssef M, Ibrahim R (2016) Molecular markers associated with high vitamin-C content in guava. J Agric Chem Biotechnol 7(3):49–55

Zambrano-Zaragoza ML, Mercado-Silva E, Ramirez-Zamorano P, Cornejo-Villegas MA, Gutiérrez-Cortez E, Quintanar-Guerrero D (2013) Use of solid lipid nanoparticles (SLNs) in edible coatings to increase guava (*Psidium guajava* L.) shelf-life. Food Res Int 51(2):946–953

Zhu X, Ouyang W, Lan Y, Xiao H, Tang L, Liu G, Feng K, Zhang L, Song M, Cao Y (2020) Anti-hyperglycemic and liver protective effects of flavonoids from *Psidium guajava* L. (guava) leaf in diabetic mice. Food Biosci 35:100574

Date Palm: Genomic Designing for Improved Nutritional Quality

Joseph Kadanthottu Sebastian, Praveen Nagella, Epsita Mukherjee, Vijayalaxmi S. Dandin, Poornananda M. Naik, S. Mohan Jain, Jameel M. Al-Khayri, and Dennis V. Johnson

Contents

J. K. Sebastian · P. Nagella
Department of Life Sciences, CHRIST (Deemed to be University), Bangalore, India
e-mail: joseph.ks@christuniversity.in; praveen.n@christuniversity.in

E. Mukherjee
Amity Institute of Biotechnology, Amity University, Noida, India

V. S. Dandin
Department of Biology, JSS College, Dharwad, India

P. M. Naik
Department of Botany, Karnatak University, Dharwad, Karnataka, India

S. M. Jain
Department of Agricultural Sciences, University of Helsinki, Helsinki, Finland
e-mail: mohan.jain@helsinki.fi

J. M. Al-Khayri (✉)
Department of Agricultural Biotechnology, College of Agriculture and Food Sciences, King Faisal University, Al-Ahsa, Saudi Arabia
e-mail: jkhayri@kfu.edu.sa

D. V. Johnson
Agriculture Consultant, Middlebrook Ave, Cincinnati, OH, USA
e-mail: djohn37@aol.com

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_43

Abstract

Date palm (*Phoenix dactylifera* L.) is one of the oldest fruit trees known where a significant amount of breeding has been carried out to improve various agronomic and nutritional characteristics. Numerous studies have been done to improve the nutritional composition and quality of the fruit due to its significant biological properties. Various strategies have been formulated for improving the agronomic characters through biofortification as well as preserving through postharvesting techniques. Modern breeding practices using molecular markers have significantly helped to identify the phenotypic, as well as genotypic, diversity for the selection of superior date palm cultivars, advanced agronomic characters like nutritional quality, disease resistance, and yield. Availability of the whole genome sequence, organellar sequence, and genetic map of date palm has helped breeders in modification and improvement of the characteristics. With the availability of bioinformatic tools and gene editing knowledge, the nutritional composition of date palm can be effectively manipulated to develop better crops along with good agronomic characters and resistance to diseases. The authors have compiled the nutritional composition of date palm fruits and detail the strategies to edit the genome and improve nutritional quality.

Keywords

Date palm · Nutritional composition · Molecular markers · Whole genome · Bioinformatics · Genetic map

1 Introduction

Date palm (*Phoenix dactylifera* L.) trees are one of the cultivated plants for centuries and its fruits served as a staple human food since time immemorial. In the world, date palm cultivation varies from region to region. There are in total more than 2000 cultivars of date growing around the world some of which are available elite varieties with good fruit qualities (Ahmad Mohd Zain et al. 2022).

Date fruits are consumed in both their fresh and dry states. Date palm fruit production volume in the world has surpassed 9.45 million mt (https://www.statista.com/statistics/960213/date-palm-market-value-worldwide/). Dried fruits of date can be stored for a long duration at ambient climatic conditions at exceptionally low cost. Date fruits have a worldwide market; primary exporters are Iran, Egypt, Saudi Arabia, Iraq, and Algeria. World-class date fruit exporters need to take some important steps with respect to good physical and chemical properties, date types, quality and size of package, main requirements of the importing countries, and transport mode (Al-Khayri et al. 2018).

As a part of traditional nutritional therapy and religious practice, date fruits have been consumed all over the world for thousands of years by millions of people. Dates are rich in carbohydrates which give instant energy to the human body. Dates are composed of vital minerals such as Fe, Mg, Ca, P, Zn, Se, K, and Mn which are important to improve their nutritional values (Aljaloud et al. 2020). Fruits contain various vitamins such as β-carotene, thiamine, ascorbic acid, riboflavin, and folic acid niacin (El Hadrami et al. 2011; Aljaloud et al. 2020). Date fruits also supply essential amino acids like histidine, valine, aspartic acid, proline, arginine, serine, methionine, threonine, lysine and alanine, glycine, leucine, tyrosine, phenylalanine, and isoleucine (Assirey 2015). Bioactive compounds are the important ingredients of date fruits including caffeic, gallic, syringic, protocatechuic, vanillic, p-coumaric, and o-coumaric acids, quercetin, kaempferol, ferulic acid, and luteolin (Vayalil 2012; Ahmad Mohd Zain et al. 2022). The superfluity of nutrition composition of date fruits makes them highly medicinal and therapeutically important crop plants.

Conventional breeding and biotechnological tools were used together to improve the quality of the date fruits, for quick growth, to develop resistance against a variety of stresses and diseases. Height reduction is one of the important objectives in the date palm breeding. In Kuwait, scientists have developed a seedless fruit by crossing tall female cvs. and a dwarf palm (*Phoenix pusilla*), combining both conventional breeding and tissue culture techniques (Sudhersan et al. 2010). Various chronic diseases of date palm reduce date fruit production. One such deadliest disease is bayoud, caused by *Fusarium oxysporum* f. sp. *albedinis*, which causes great losses in Morocco and Algeria (Sedra and Lazrek 2011). To control the disease, various control measures were applied including chemical and biological applications, but among them all, developing resistant varieties is the most promising way to fight against the bayoud disease. Date palm cv. Taqerbucht from the oases of Salah and Tidikelt, Algeria, is found to be the best resistant cultivar against bayoud (Bouguedoura et al. 2015). In Tunisia, a new disease named brittle leaf disease has destroyed around 40,000 date trees in a short time span due to leaf wilting; the cause of the disease is still unknown (Triki et al. 2003). The date palm crop has a breeding limitation due to its growing environmental conditions where the plant must survive extreme salt and drought stress. Another problem is sex determination; until the onset of fruiting, it is not possible to identify the sex of the plant which matures at 5–7 years of age. Farmers are using traditional and rudimentary agricultural techniques from past generations. Water scarcity, a primitive level of farm and crop

management, applying traditional methods of irrigation, limited use of advanced breeding tools, lack of skilled labor, and a low level of mechanization are some constraints to date palm breeding (Al-Khayri et al. 2015).

2 Chemical Composition and Biochemical Pathways

2.1 Chemical Composition

The plant species of the family Arecaceae, *Phoenix dactylifera*, has been a staple food in the Indus valley region and Middle East since before than sixth millennium BCE. Extensive research has been conducted on its nutritional composition and related molecular biochemistry. The overall chemical composition of date palms suggests that it consists of a very large number of nutrients and associated molecules. Due to this long list of nutritional chemicals, the benefits of date palm consumption are many. Starting from its activity as an energy booster, it is also credited with elevating the iron content in the body. It helps in digestion and has various pharmacological activities. Due to their appropriate balance in composition, they have also been regarded as being sufficient for consumption during fasting. All these activities are credited with the presence of innumerable chemicals in the plant. The overall list of contents in various species of date palms is comprehensively reviewed in the following section.

Almost every organ of the plant is rich in a variety of nutritional compositions. The stem of date palm consists of important nutritional compounds such as stilbene (Fernández et al. 1983a). Along with this, the same part of the plant also contains important steroid, polysterol, and stigmastane molecules (Fernández et al. 1983b). In addition, the stems also possess sterols such as ergosterol along with aromatic carboxylic acids like benzoic acid and its derivative like 3,5-dihydroxy-4-methoxy benzoic acid (Fernández et al. 1983a). Triterpenoids have already been tested and found to be very crucial anti-inflammatory, anti-pyretic, and hepatoprotective agents in medicine. For conditions associated with weight loss and reduction of cholesterol levels, it is also highly necessary to reduce the risk of heart attacks which are aided by the presence of plant sterols like campesterol.

The seeds of date palm plants also consist of a variety of important phytochemicals that boost the immunological capabilities of humans. The long list starts with the presence of amino acids like alanine in seeds which help in metabolic activities in the human body (Hussein and El Zeid 1975). It relates to the muscle and improvement of physical functioning in humans. Notably, the concentration of amino acids is higher in the unripe stages of the seeds as compared to the ripe ones, as arginine, glutamic acid, aspartic acid, leucine, lysine, and serine were the highest in unripe stages, while proline and glycine were the highest in ripe stages (Auda et al. 1976). The seeds of the date palm are home to mostly several SFAs (Jindal and Mukherjee 1970; Mossa et al. 1986) (Table 1, Fig. 2). Pivotal agents to nutrition include the presence of saccharides obtained from plants as they help magnify nutrition levels, prevent blood clotting, and are one of the major products consumed as staple food in

Table 1 Chemical composition and health benefits of date palm

Plant Part	Chemical	Specification	Health benefits	Reference
Stem	*Cis*-3,5,3′,5′-tetrahydroxy-4-methoxystilbene	Stilbene (olefin)	Antioxidant benefits	Fayadh and Al-Showiman 1990
	Trans-3,5,3′,5′-tetrahydroxy-4-methoxystilbene	Stilbene (olefin)	Anticancer therapy	Fayadh and Al-Showiman 1990
	Trans-3,5,4′-trihydroxystilbene	Stilbene (olefin)	Phytoalexin Anticancer therapy	Fernández et al. 1983
	5α-Stigmast-22-en-3,6-dione	Steroid	Anti-inflammatory	Fernández et al. 1983
	5α-Campestan-3,6-dione	Steroid	Phytosterol, lowers cholesterol in the gut	Fernández et al. 1983
	Stigmasta-4-ene-3-one	Phytosterol	Hypoglycemic metabolite	Toghueo 2019
	Stigmasta-4,22-diene-3-one	Stigmastane	Potential biomarker	Liolios et al. 2009
	Stigmasta-4,22-diene-3,6-dione	Stigmastane	Antiplatelet aggregation	Fernández et al. 1983
	Lupeol	Pentacyclic triterpenoid	Anti-inflammatory drug, acne treatment	Alam et al. 2009
	Lupeol acetate	Pentacyclic triterpenoid	Anti-arthritic, leishmanicidal, anti-inflammatory, antifungal activity	El-Far et al. 2019
	Hydroxystigmast-4-en-3-one	Stigmastane (oligopeptide)	Reduction of blood cholesterol	Fernández et al. 1983
	Hydroxystigmasta-4,22-dien-3-one	Stigmastane (oligopeptide)	Antiparasitic, radical scavenging capacity	Fernández et al. 1983
	6β-Hydroxy-campest-4-ene-3-one	Stigmastane (oligopeptide)	Cytotoxicity against lung cancer cell lines	Fernández et al. 1983
	Campest-4-en-3-one	Ergosterol	Treatment of cardiovascular diseases	Fernández et al. 1983
	Campest-4-en-3,6-dione	Ergosterol	Promotes energy consumption, reduces visceral fat	Suzuki et al. 2007
	Campest-4-en-3-one	Ergosterol	Lowers LDL and cholesterol	Suzuki et al. 2007
	Benzoic acid	Aromatic carboxylic acid	Antimicrobial food preservative	Das et al. 2017

(continued)

Table 1 (continued)

Plant Part	Chemical	Specification	Health benefits	Reference
Seeds	Alanine	Amino acid	Hypoglycemia, urea cycle disorder treatment	El-Sohaimy and Hafez 2010
	Arginine	Amino acid	Chest pain, pregnancy complications	El-Sohaimy and Hafez 2010
	Glutamic acid	Amino acid	Excitatory neurotransmitter in central nervous system	Eeuwens 1978
	Aspartic acid	Amino acid	Fatigue, athletic performance, muscle strength	Assirey 2015
	Leucine	Amino acid	Energy for skeletal muscle	Shaba et al. 2015
	Lysine	Amino acid	Proper growth and carnitine production	Shaba et al. 2015
	Serine	Amino acid	Enhancement of memory and thinking skills	Abdalla et al. 2020
	Proline	Amino acid	Collagen producer	Dhawi and Al-Khayri 2008
	Glycine	Amino acid	A component of creatine and protection against muscle loss	Dhawi and Al-Khayri 2008
	Arachidic acid	Saturated fatty acid	Chemical messenger released by muscles, detergent production	Abdalla et al. 2020
	Tricosanoic acid	Saturated fatty acid	An internal standard, hair growth stimulant	Al-Shahib and Marshall 2003a
	Stearic acid	Saturated fatty acid	Emulsifier, emollient, cosmetic product preparation, rubber processing	Nehdi et al. 2018
	Palmitoleic acid	Saturated fatty acid	Anti-inflammatory, improves insulin sensitivity	Al-Shahib and Marshall 2003b
	Palmitic acid	Saturated fatty acid	Industrial mold release agent production, soap, detergent production	Saafi et al. 2008

(continued)

Table 1 (continued)

Plant Part	Chemical	Specification	Health benefits	Reference
	Oleic acid	Unsaturated fatty acid	Inflammation reduction, cholesterol reduction	Al-Shahib and Marshall 2003b
	Myristic acid	Saturated fatty acid	Flavor ingredient, soap or detergent production, surfactant, cleansing agent	Shehzad et al. 2021
	Margaric acid	Saturated fatty acid	Anti-tumor, glutamate and lipid metabolite in vivo	Abdalla et al. 2020
	Linolenic acid	Saturated fatty acid	Cholesterol and high blood pressure reduction, atherosclerosis treatment	Attia et al. 2021
	Behenic acid	Saturated fatty acid	Lubricating oils, anti-foaming agents in detergents	Soliman et al. 2015
	Lauric acid	Saturated fatty acid	Treatment of viral infections, soap and detergent production	Nehdi et al. 2018
	Heneicosanoic acid	Saturated fatty acid	Foams, paints, viscous material production in industries	Di Cagno et al. 2017
	Linoleic acid	Saturated fatty acid	Body-building, fitness, weight loss, cosmetics, and personal care products production	Al Juhaimi et al. 2020
	Rhamnose	Saccharide	Antiaging, anti-wrinkle, natural furanose production	Khallouki et al. 2018
	Fructose	Saccharide	Manufacturing low-calorie products, flavored water, taste enhancer	Nadeem et al. 2019
	D-galactose	Saccharide	Energy source, treatment of hepatitis C, hepatic cancer, precursor to glucose production	Ataei et al. 2020

(continued)

Table 1 (continued)

Plant Part	Chemical	Specification	Health benefits	Reference
	Glucose	Saccharide	Treatment of hypoglycemia, source of energy for the brain, overall health maintenance	Siddiqi et al. 2020
	Lactose	Saccharide	Ethanol production, pharmaceutical industry, diluent, bulking agent	El-Kholy et al. 2019
	D-mannose	Saccharide	Treatment of carbohydrate-deficient glycoprotein syndrome, alternative to antibiotics	Alyileili et al. 2020
	Xylose	Saccharide	Food sweetener, flavoring agent, diagnostic agent for malabsorption	Ataei et al. 2020
	Zinc	Essential mineral	Metabolism functioning, developing immune system, improved sense of taste and smell, wound healing	Adenekan et al. 2018
	Iron	Essential mineral	Essential for blood production, healthy skin, bones, hair, immune system booster	Alem et al. 2017
	Manganese	Essential mineral	Metabolism of various amino acids, carbohydrate, bone formation, anti-inflammatory, blood clotting	Kuhiyop et al. 2020
	Aluminum	Essential mineral	Food conservation, increase effects of vaccines and medicines in the body	Badarusham et al. 2019
	Potassium	Essential mineral	Increased uptake of nutrients into the body, contraction of muscles, nerve functioning	Alem et al. 2017

(continued)

Table 1 (continued)

Plant Part	Chemical	Specification	Health benefits	Reference
	Phosphorus	Essential mineral	Aid in ATP synthesis, production of proteins for growth, helps build bones	Bijami et al. 2020
Pollen	β-Sitosterol	Phytosterol	Benign prostatic hyperplasia treatment, cholesterol reduction	Al-Samarai et al. 2018
	Rutin	Flavonoid	Antioxidant, allergy treatment, vitamin C effect enhancer, arthritis treatment	El-Kholy et al. 2019
	Quercetin	Flavonoid	Heart disease risk reduction, cancer treatment, allergy treatment, infection reduction	Otify et al. 2019
	β-Amyrin	Pentacyclic triterpenoid	Potential antinociceptive, gastroprotectant, hepatoprotectant, antioxidant, anti-inflammatory	Hamed et al. 2017
	Estrone	Steroid	Perimenopausal and postmenopausal symptom treatment	El-Sisy et al. 2018
	Estrogen	Steroid	Regulation of female reproductive system, bone development, treatment of menopausal symptoms	Otify et al. 2021
Fruits	Alanine	Amino acid	Hypoglycemia, urea cycle disorder treatment	Mohammadi et al. 2018
	Valine	Amino acid	Energy enhancer, muscle growth, tissue repair	Kadum et al. 2019
	Tyrosine	Amino acid	Improvement in memory, attention, and focus	Abdul-Hamid et al. 2019
	Tryptophan	Amino acid	Melatonin, serotonin, and niacin producer	Yaish et al. 2017

(continued)

Table 1 (continued)

Plant Part	Chemical	Specification	Health benefits	Reference
	Threonine	Amino acid	Treatment of nervous system disorders	Assirey 2015
	Serine	Amino acid	Enhancement of memory and thinking skills	Assirey 2015
	Proline	Amino acid	Collagen producer	Al-Khayri and Al-Bahrany 2004
	Phenylalanine	Amino acid	Treatment of Parkinson's disease, depression, and other nervous disorders	Assirey 2015
	Methionine	Amino acid	Treatment of liver disorders, helps in cell cycle, wound healer	Assirey 2015
	Lysine	Amino acid	Improvement in calcium uptake and reduction in anxiety and stress, treatment of cold sores	Assirey 2015
	Leucine	Amino acid	Provide ATP during exercise, tissue regeneration, protein synthesis, and metabolism	Assirey 2015
	Isoleucine	Amino acid	Blood sugar regulation, hemoglobin synthesis, energy level maintenance	Assirey 2015
	Histidine	Amino acid	Treatment of allergic diseases, kidney failure, ulcers, rheumatoid arthritis	Assirey 2015
	Glycine	Amino acid	Treatment of benign prostatic hyperplasia, stroke, schizophrenia, protection of kidneys	Assirey 2015
	Glutamic acid	Amino acid	Protein production in the body, transforms into glutamate	Al-Shahib and Marshall 2003b

(continued)

Table 1 (continued)

Plant Part	Chemical	Specification	Health benefits	Reference
	Glutamine	Amino acid	Helps gut function, immune system, provides ATP to the body	Zouine and El Hadrami 2007
	Asparagine	Amino acid	Treatment of imbalanced foods, production of proteins, functioning of nervous systems	Abbas and Saad 2009
	Aspartic acid	Amino acid	Boosts up low testosterone levels, protein biosynthesis	Assirey 2015
	Arginine	Amino acid	Chest pain, pregnancy complications	Sghaier et al. 2009
	Fructose	Saccharide	Manufacturing low-calorie products, flavored water, taste enhancer	Assirey 2015
	Sucrose	Saccharide	Syrup processing, sweet drink softener, detergent, emulsifier carrier	Nwanekezi et al. 2015
	Glucose	Saccharide	Treatment of hypoglycemia, source of energy for the brain, overall health maintenance	Assirey 2015
	Calcium	Essential mineral	Movement of muscles, transfer of essential messages across neurons	Al-Shahib and Marshall 2003b
	Zinc	Essential mineral	Metabolism functioning, developing immune system, improved sense of taste and smell, wound healing	Al-Shahib and Marshall 2003b
	Potassium	Essential mineral	Increased uptake of nutrients into the body, contraction of muscles, nerve functioning	Uddin and Nuri 2021
	Phosphorus	Essential mineral	Aid in ATP synthesis, production of proteins for growth, helps build bones	Awofadeju et al. 2021

(continued)

Table 1 (continued)

Plant Part	Chemical	Specification	Health benefits	Reference
	Manganese	Essential mineral	Metabolism of various amino acids, carbohydrate, bone formation, anti-inflammatory, blood clotting	Ibraheem 2021
	Magnesium	Essential mineral	Energy promotion, maintenance of normal nerve functioning, immune system	Uddin and Nuri 2021
	Iron	Essential mineral	Essential for blood production and healthy skin, bones, and hair, immune system booster	Lamia and Mukti 2021
	Ascorbic acid	Vitamin C additive	Collagen formation, wound healer, cartilage, teeth, bones, repair of body tissues	Rajan et al. 2021
	Iso-chlorogenic acid	Phenolic acid	Lowers blood pressure, blood sugar, and weight	El-Far et al. 2016
	γ-Aminobutyric acid (GABA)	Neurotransmitter	Brain neurotransmitter, anxiety reducer, pain reliever	López-Córdoba 2021
	Vitamin A	Vitamin	Normal vision, reproductive ability, helps organs work well	Hinkaew et al. 2021
	Thiamine	Vitamin additive	Pyruvate metabolism, conversion of carbohydrates to energy, nerve signal transmission	Shehzad et al. 2021
	Riboflavin	Vitamin additive	Fat and protein metabolism, conversion of carbohydrates to energy	Awofadeju et al. 2021
	Nicotinic acid	Vitamin additive	Triglyceride and cholesterol improvement	Tawfek et al. 2021

our diets (Noorbakhsh and Khorasgani 2022). As far as seeds of date palm are concerned, they contain a large number of saccharides, essential for human life such as rhamnose, xylose, D-mannose, lactose, glucose, D-galactose, and most importantly, the sweetest sugar, fructose (Jindal and Mukherjee 1969; Mahran et al. 1976; Al-Whaibi et al. 1985). Zinc, phosphorus, potassium, aluminum, manganese, and iron are a few of the beneficial elements found along with other chemicals in the seeds (Sawaya et al. 1982; El-Shurafa et al. 1982; Hamad et al. 1983).

The pollen of date palms also has been found to contain a few important biomolecules like β-sitosterol, which resembles cholesterol molecules and helps to reduce hypercholesterolemia (Mahran et al. 1976; Kikuchi and Miki 1978; Fernández et al. 1983b). Another fascinating glycoside, formed by the combination of flavonol quercetin and disaccharide rutinose, rutin is found in the pollen of the plant (Mahran et al. 1976). It is chiefly used in the production of medicines which has a similar function as that of sitosterol, in lowering cholesterol levels. In addition, it is also used to reduce arthritis pain and prevent blood clotting. The flavonoid quercetin is also present in the pollen along with female sexual hormones like estrogen and estrone (Hassan and Abou El Wafa 1947; Bennett et al. 1966; Mahran et al. 1976; Mossa et al. 1986). Estrone in the body is converted to estrogen when needed; thus, the former serves as a repository for hormone estrogen in females. A triterpenoid, i.e., β-amyrin, is also found in pollen that is responsible for the improvement of glucose tolerance owing to its anti-inflammatory and antioxidant effects (Mahran et al. 1976).

The most important and maximum number of phytochemicals exists in the edible part – "fruit." Nearly all amino acids like alanine, arginine, etc. and saccharides like fructose, glucose, and sucrose (Auda et al. 1973; Hussein and El Zeid 1975; Auda et al. 1976; Sawaya et al. 1983b) are present. Elements like Ca, Fe, Mg, Mn, P, K, and Zn with vitamin additives like ascorbic acid, nicotinic acid, riboflavin, thiamine, and vitamin A are also found (El-Shurafa et al. 1982; Sawaya et al. 1983; Mossa et al. 1986). Natural products with pharmaceutical potentials like neurotransmitter γ-aminobutyric acid (GABA) and iso-chlorogenic acid are added compositions (Cadi et al. 2021; Hattori et al. 2021) (Table 1).

2.2 Chemical Type, Structure, and Biochemical Pathways of Production

Date palms have a plethora of important nutrients and chemicals in their composition ranging from vitamins to amino acids and fatty acids. The extensive details about all the parts of the date palm containing the respective nutrients have already been explained. Specifically, when the leaf fiber extracts of date palm were studied, constituents obtained in decreasing order of their percentages were α-cellulose (58%) to pectin (2.3%) (Pandey and Ghosh 1995). In fact, a good replacement for synthetic fibers exists through the microcrystalline cellulose extracted from the fruit fibers of the date palm. Lignin and hemicellulose were also isolated from the date palm fiber, indicating their richness in nutraceutical properties (Hachaichi et al. 2021).

The list of important therapeutic and immune condition-boosting biomolecules present in the fruits of the date palm is long, as already mentioned. The range commences at amino acids like alanine and terminates at nicotinic acid and other vitamin additives. Amino acids are structurally identical, except for the -R group which determines the nature of the molecule (Figs. 1 and 2). They are zwitter ionic

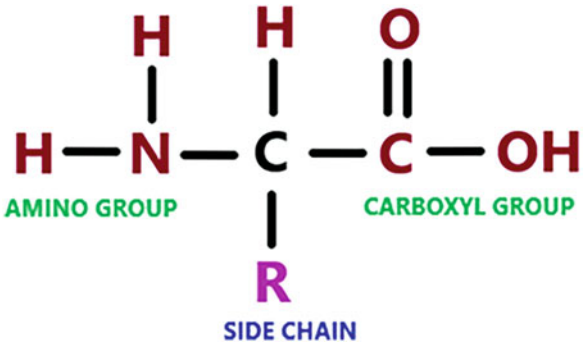


Fig. 1 General structure of amino acids

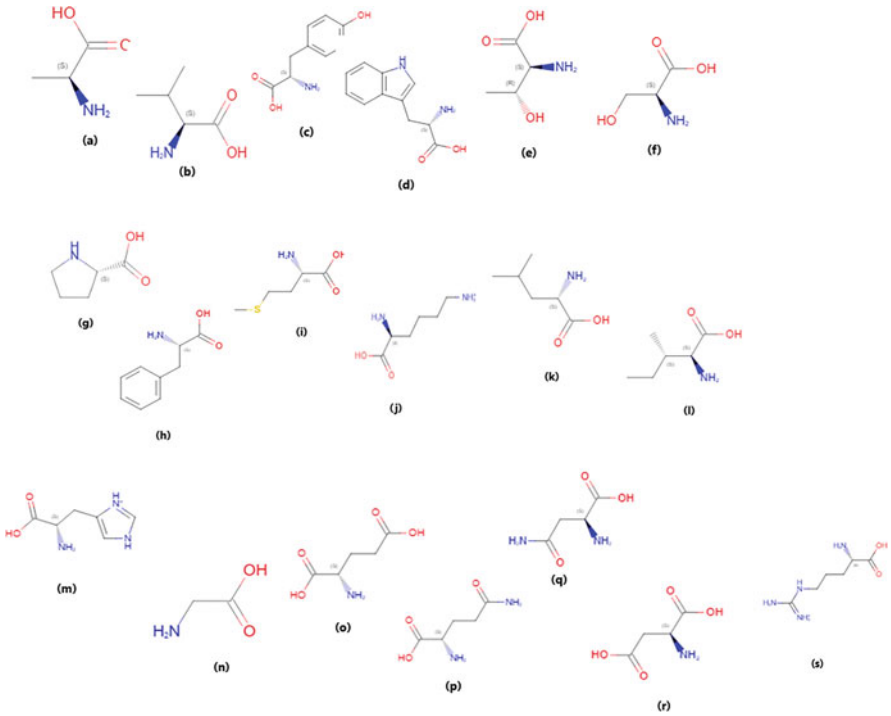


Fig. 2 Structures of a few amino acids found in the fruits of date palm. (Source: PubChem) [a-Alanine, b- Valine, c- Tyrosine, d- Tryptophan, e- Threonine, f- Serine, g- Proline, h- Phenylalanine, i- Methionine, j- Lysine, k- Leucine, l- Isoleucine, m- Histidine, n- Glycine, o- Glutamic acid, p- Glutamine, q- Asparagine, r- Aspartic acid, s- Arginine]

molecules that exhibit the acidic and basic character simultaneously at a given set of environmental conditions.

Other parts of the plant are also rich in several important pharmaceutical biomolecules and guarantee a healthy immune system when consumed on a regular basis. One of the major reasons why date palms are included in our diet during a very physically demanding fast is because of the availability of enormous amounts of carbohydrates and fats, which is also the reason why date palms exist under "Ojas" or energy-giving food in Ayurveda (Haas 2015). Moreover, high amounts of soluble fibers (Maou et al. 2021) and the alkaline nature of fresh date palms (Alghamdi et al. 2019) further increase the beneficial properties of its consumption because it increases colon performance and compensates for the increased stomach acidity, respectively, after fasting. The "sattvic" nature of soft and succulent fresh dates as described in Ayurveda aids in cooling down the heated body after a long fast. Other important biomolecules present in date palms are stilbenes (Fig. 3a), steroids (Fig. 3b), phytosterols (Fig. 3c), minerals like calcium (Fig. 3d), saccharides (Fig. 3e), saturated (Fig. 3f) and unsaturated fatty acids (Fig. 3g), vitamins (Fig. 3h), triterpenoids (Fig. 3i), flavonoids (Fig. 3j), and carboxylic acids (Fig. 3k), as listed in Table 1.

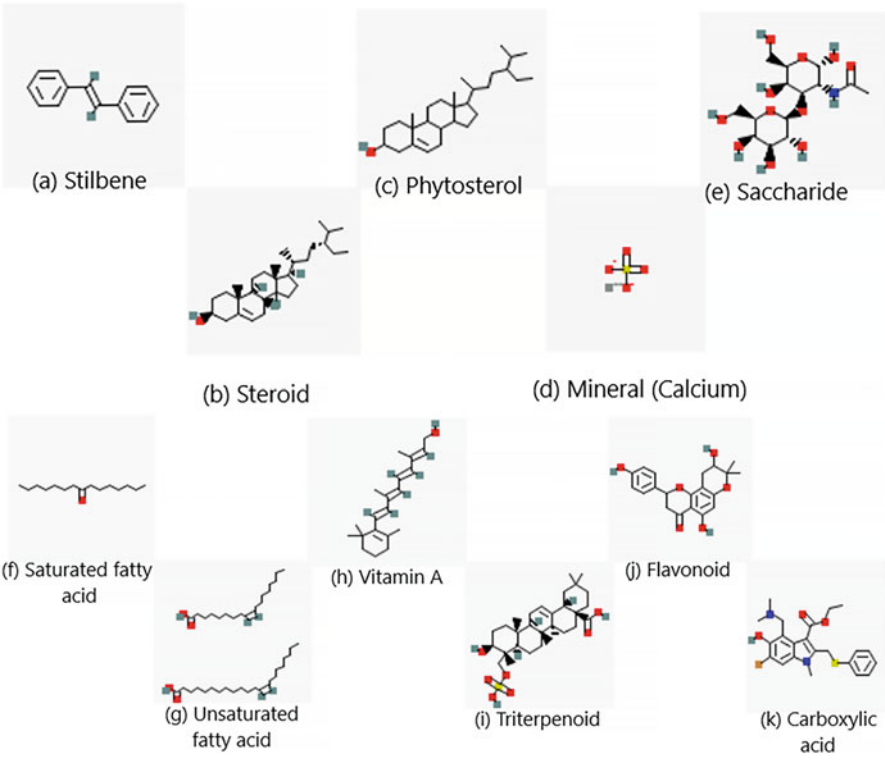


Fig. 3 Structure of a molecule from each group of biomolecules present in different parts of date palm. (Source: PubChem)

The biochemical pathways to produce biomolecules generally follow a distinct cascade of reactions in the organism. In the case of date palms, there are a wide range of biomolecules present in almost all organs as mentioned. When focusing on stems, olefins are present in the majority, which are produced by a variety of reactions like hydrogenation, dehydration, hydro-deoxygenation, or steam cracking. It has been found that almost all the important olefins are ultimately produced from the intermediates obtained from the basic metabolic cycles in respiratory pathways such as glycolysis or citric acid cycle. An example of such a pathway has been found starting from glucose to olefins like stilbene and various other secondary metabolites in plants (Fig. 4).

Very interestingly, it has been observed that there are different changes in the amount and number of metabolites in the context of amino acid biosynthesis in date palms when subjected to UV-B light irradiation (Agwa 2021). There is a cascade of reactions that give rise to multiple metabolites in between two steps that can be significant to humankind, both in terms of economy and nutrition. The saccharide biosynthesis in date palms occurs through another series of reactions and biochemical pathways, as in any other fruit species as depicted by researchers already.

Not only date palms have the biochemical pathways to produce metabolites, but they have been studied in other species, such as in maize. Sharma et al. (2012)

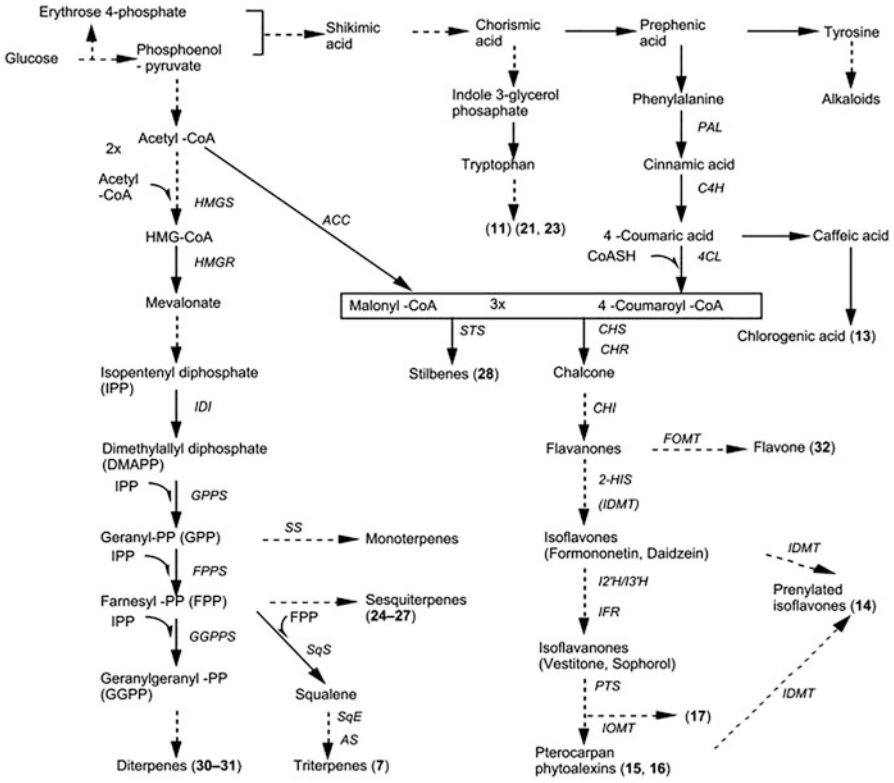


Fig. 4 Biosynthesis of various olefins in plants by metabolic pathways. (Modified from Bacigalupa et al. 2018)

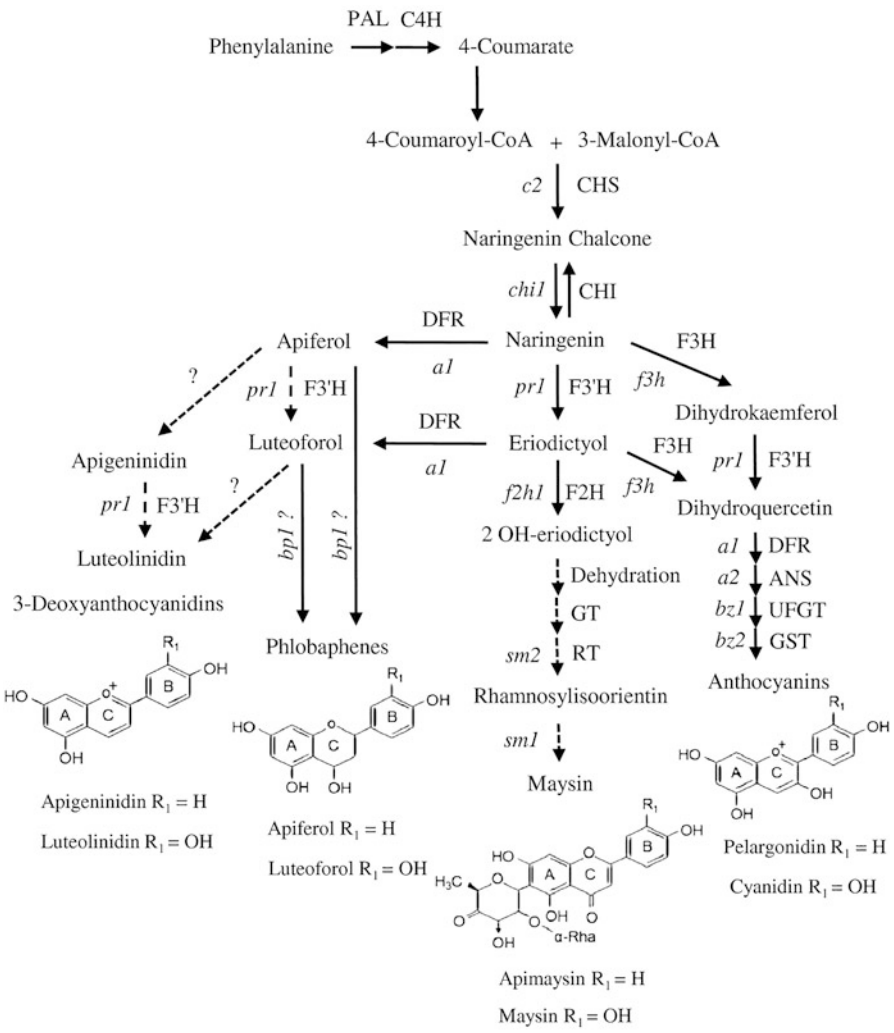


Fig. 5 Flavonoids' biosynthetic pathway in maize (Sharma et al. 2012)

reported on flavonoid biosynthesis in maize (Fig. 5). Identification of crucial genes responsible for the authentic metabolism of raw products is important. In date palms, extensive research is ongoing on its biochemical pathways of metabolite production.

2.3 Medicinal/Physiological Properties and Functions in Relation to Human Health

Date palms are also consumed by humans because of their restorative and curative properties. The medicinal properties of the plant range from the simplest of

Fig. 6 Protective activities of *Phoenix dactylifera* against various ailments

symptoms like cold and fever to life-threatening diseases such as liver cancer and bronchial catarrh (Fig. 6). The credit for being a remedy to this long list of ailments goes to the enormous number of phytochemicals present in the plant. The main biomolecules present in the plant are vitamins, minerals, flavonoids, carbohydrates, steroids, alkaloids, and many others (Ahmad Mohd Zain et al. 2022). Some of the important medicinal properties of the date palm are listed in Table 1.

2.3.1 Anticancer Activity

One of the leading causes of death today is cancer and an extensive number of research is being done to prevent and treat this disease. The plant under study contains a good number of anticancerous agents and studying the effects of those molecules in rat models can increase its use for therapeutic purposes. For the prevention of hepatocellular carcinoma, restoration of antioxidant enzymes in the liver to accurate levels is very important, which can be performed using biomolecules like quercetin, luteolin, or glucans already existent in date palm (Lamia and

Mukti 2021). For mammary cancer in humans, the effect of various concentrations of water extracts of Jordanian dates was assessed on MCF-7 breast cancer cell lines. On analyzing the effect of anticancerous agents using the MTT assay and calculating the percentage of inhibition, it was found that 100 mg/ml was the most effective in exhibiting the maximum inhibition (Al-Sayyed et al. 2021). Acetone extracts of date fruit (Ajwa variety) demonstrated cytotoxic effect on oral squamous cell carcinoma cell lines by inhibiting their growth and proliferation and inducing cell death by apoptosis (Shahbaz et al. 2022). Apart from these, the consumption of dates also increases the health condition of colon cells by colon cancer cell growth inhibition and enhancing beneficial bacterial growth in the colon (Eid et al. 2014). Various extracts of ethanol, acetic acid, and acetone using date palm have also reduced the proliferation of pancreatic stellate cells (PSC) in vitro. In addition, it also reversed the fibrotic phenotype of PSCs, thereby reducing fibrosis (Al Alawi et al. 2020). Thus, on further research, there is a high capacity for date palm to serve as a potential anticancerous agent, and isolation of the novel phytochemicals responsible for the same can be done on a commercialized scale for treatment purposes.

2.3.2 Anti-diarrheal Activity

Diarrhea can be acute or chronic depending on the immune system of the individual. When castor oil was used to induce diarrhea in Wistar rats, it was observed that extracts of date palm used in the concentration of 1000 and 1500 mg/kg of body weight were successful in reducing the intensity of diarrhea (Agbon et al. 2013). The results were almost the same in another experiment when loperamide was administered as a control agent (Megbo et al. 2017). Comparatively, date palm aqueous extracts possessed completely nontoxic anti-diarrheal activity when used on male Wistar rats. Phytochemical screening and spectrophotometric methods revealed the high crude ash, carbohydrate contents with saponins, tannins, and flavonoids along with a low anti-nutrient concentration in date palm extracts.

2.3.3 Anti-ulcer Activity

When gastric ulcer was introduced in rats using ethanol, different extracts of both dialyzed and undialyzed date fruits were administered to check the effect of dates on gastric ulcers in humans, at 4 ml/kg for 14 days. It was observed that both aqueous and ethanolic extracts could work successfully toward the reduction of severe gastric ulcers in rats (Al-Qarawi et al. 2005). The presence of the ethanolic extract of undialyzed date palm fruit could reduce the level of gastric mucin, histamine, and gastrin in the rats. The effect of date palms on peptic ulcer conditions was further affirmed when the report of Gangwar et al. (2014) suggested that chloroform extract of its leaves contained gastroprotective activity. This was supported by the inhibition of ulcer index, mean ulcer score, and reduction in gastric content when the pH increased and acidity levels, both free and total, decreased, along with doses. Thus, it has already been proven that painful sores can easily be removed by the consumption of date fruits, which become hard to remove even sometimes using synthetic molecules.

2.3.4 Anti-hepatotoxic Activity

As of 2021, diabetes ranks as one of the major reasons of the death of humans in the world due to a high cholesterol diet and eventually brings with it associated disorders. Liver health deterioration is a concern for people, irrespective of age, which can further be responsible for chronic illnesses, and thus, treatments for the prevention of this disease are the need of the hour. It has been observed that when date palm's hydroalcoholic extracts were administered to rats having hepatotoxic effects due to diabetes, comparatively healthier sinusoids and hepatocytes were found (Zheng et al. 2021). On the other hand, when liver diseases were untreated, severe histopathological conditions such as dilated portal veins, fatty degeneration, and necrotic nuclei were seen. Since this observation was a remarkable advancement as conducted in 32 adult rats, further clinical research is solicited for the execution of the hepatoprotective activity of date palms on humans. In addition, Onyilo et al. (2021) reported the hepatoprotective potential of date palm's aqueous extract when fat-induced damaged liver cells in adult Wistar rats were found to be reversed and Kupffer cells activated with changes in the level of hepatic enzymes as well as hepatocyte configuration.

2.3.5 Antioxidant Activity

Antioxidants are important chemicals needed to protect the body against free radicals produced by catabolism and capable of causing cancer (Moslemi et al. 2022). They are necessary substances in the body and have been found in great abundance in date palms (Zihad et al. 2021). High quantities of phenolic, tannin, and flavonoid contents were recorded in three varieties of date palms, namely, Ajwah, Sakkari, and Safawy, using UPLC-QTOF-MS assays. Radical scavenging activity, both DPPH and hydroxyl, was found in all of them along with a total antioxidant capacity of IC_{50} 87–192 μg/ml. Results were obtained, almost along the same lines when antioxidant and total phenolic levels were compared among methanolic extracts of four different date palm varieties (Bensaci et al. 2021). Folin-Ciocalteu and DPPH test along with superoxide anion and reducing power test exhibited the highest antioxidant levels of 0.006 mg/ml in the Chtaya variety using the cyclic voltammetry method. Thus, consumption of the fruit can also aid in keeping deadly diseases like cancer at bay, as proved in the literature already.

2.3.6 Anti-inflammatory Activity

Out of the many visible therapeutic activities of date palms, one of the most important is its anti-inflammatory effect. In a report by Barakat et al. (2020), it has been proven that date palm seeds could exhibit anti-inflammatory effects on inflammation induced by LPS in RAW 264.7 cells. This was mainly done against the nitric oxide release and subsequent iNOS protein expression. In addition, suspensions of date palm pollen have been shown to improve immune-histochemical and histopathological conditions in rats with hyperplasia (Elberry et al. 2011). The presence of inflammation was confirmed due to the presence of chemical messengers like IL-6, IL-8, TNF-α, and IGF-1. Along with upregulation of autocrine or paracrine receptors, the date palm has been shown to modulate cytokine expression through a

protective effect by suppressing inflammation. Cellular proliferation was accompanied by a reduction in apoptosis in the ventral prostate of rats with hyperplasia on the administration of date palm.

2.3.7 Antimicrobial Activity

The term "viral infection" needs no new introduction in the present-day scenario of 2021 where the entire world fought and is fighting against the deadly coronavirus pandemic. At this juncture, it is highly essential that the availability of the number of antiviral commodities is increased as much as possible. For the same, the leaf extract of date palm had been tested in vitro using an MTT assay on the embryonic fibroblast cells of NIH/3 T3 mouse. Takdehghan et al. (2021) reported a significant decrease in CRP levels as well as elevation in partial pressure of oxygen levels in those patients administered with date palm leaves five times daily along with the recommended dosage of medicines as compared to the patients strictly on routine medication. When checked on days 7 and 14 post-administration, the inference was drawn and thus was recommended as a safe add-on along with the daily dosage of treatment. The efficiency of date palms against bacteria, both Gram-positive and Gram-negative, has also been studied when Ghosh (2021) reported that methanolic extracts of its seeds have exhibited antibacterial activities against *Klebsiella pneumonia* and *Escherichia coli*. Along with that, it has also reduced the side effect on the testosterone and muscles of rats and improved other hormonal functions.

2.3.8 Antihyperlipidemic Activity

Due to the possibility of developing a heart disease, it is necessary to maintain a low blood lipid level or triglyceride. Lipoproteins are essential for providing energy in the body but only up to a certain level, because beyond that level, the severity of medical conditions can escalate to diabetes, alcoholism, kidney disease, and other chronic cases too. Date palm possesses antihyperlipidemic activity as has been reported by Innih and Lorliam (2021) in an experiment where it was determined whether date palm has hyperlipidemic properties, owing to the phytochemicals present in the plant. Against hyperlipidemia induced by margarine in rats, the differential blood count and lipid profile analysis revealed a decrease in cholesterol and LDL levels whenever date palms were used as a part of the therapeutic routine. The results were further confirmed when Silabdi et al. (2021) explained the reduction in total and LDL cholesterol along with that in triacylglycerol while the level of HDL in the blood of rats increased. These results were obtained from three different varieties of Algerian date palms, namely, Deglet Nour, Ghars, and Degle Baida. Date palms play a crucial role in keeping lipid and other lipid-related levels such as SGPT, ALT, LDH, and GOT at a normal level without any other side effects compared with synthetically derived products.

2.3.9 Anti-nephrotoxic Activity

To keep one ailment at bay, the chemical medications may invite other illnesses. For severe diseases like lymphoma or leukemia, doxorubicin is a widely used drug with a high therapeutic index, linked with causing nephrotoxicity in patients as a side effect.

Wang et al. (2019) reported the activity of date palm extract on the nephrotoxic effects caused by doxorubicin in patients. Levels of cardiac markers like LDH-1 were seen to decrease and nephrotoxicity was seen to be reversed on the administration of date palm extract. As far as nephroprotection is concerned, there was amelioration in the urine flow rate and a significant increase in the urine sodium-potassium ratio due to the activation of Na^+/K^+ ATPase in the presence of date palm extract. In addition, there can also be an improvement in the functioning of nephrons using its antioxidant potential on proximal tubular damage. On further investigation, the reason for the observed nephroprotective activity of date palm alluded to elevation in the level of antioxidant enzymes which possessed the capability of free radical scavenging (Al-Qarawi et al. 2008). Dates were already known to act as strong antioxidants for their mechanism of counteraction of free radicals. Hence, nephroprotection is a major function of date palms owing to their radical scavenging capacity.

2.3.10 Antimutagenic Activity

The presence of any external agent or any other factor might be responsible for transformations in organisms, which can even be fatal in some cases and are phenotypically invisible. To safeguard or inhibit such a process from occurring, it is crucial that an intensive number of research is done on the naturally present antimutagenic substances. Vayalil (2002) reported the antimutagenic property of date palms on *Salmonella* tester strains TA-98 and TA-100. The property was tested for His^+ revertant formation in the strains which exhibited 50% dose-dependent inhibition of benzo(α)pyrene-induced mutagenicity on the administration of date palms. Not only in this case but date palm's antimutagenicity was also examined against N-nitroso-N-methylurea by the activity of chromosome aberration and other DNA fragmentation assays. On administration of date palm extract, it was seen that the DNA damage caused by the mutagen was restored back to its normal conditions (Diab and Aboul-Ela 2012). This was proven by the reduction in the chromosomal aberration and micronuclei in the bone marrow, and no further DNA fragmentation in hepatic cells was seen when date palm extracts were introduced before and after the treatment by intraperitoneal injection in mice. Furthermore, research on the medicinal properties of date palm might open doors to more innovative and therapeutic discoveries helpful to the economy and human health.

2.4 Methods of Biofortification: Agronomic and Postharvest Techniques

2.4.1 Agronomic Biofortification

The procedure of application to increase the nutritional content in plants during their growth phase naturally through agronomic methods is termed biofortification. This is essential to maximize the mineral uptake by plants or nutrient availability while consuming the fruits and can be done by application of several mineral fertilizers to the plant parts or soil. To date, there have been numerous trials to improvise the mineral and nutrient values in several species of plants, but there is also huge

potential for the same in the remaining plants (Stangoulis and Knez 2022). In date palm too, efforts were made to increase its nutritional value. However, the level of research conducted is very limited and needs more attention. With the advent of newer strategies and scientific advancements like insertional, somaclonal, and site-directed mutations, biofortification in plants has become easy. Along with that, provisions for targeted and more accurate mutagenesis have also increased with the discovery of CRISPR-Cas gene editing and base editing systems in engineering date palm genomes (Sattar et al. 2017).

There are several issues associated with breeding a plant, including both biotic and abiotic factors. Abiotic components include the issues of dry soil, inadequate mineral contents, scarcity of water, or insufficient root activity in a few cases. All of them can be improved by advanced techniques of agronomic biofortification like the usage of NPK, potassium sulfate, nitrate, silicate, iron sulfate, urea, zinc sulfate, etc. nutrients externally. Numerous studies have been conducted on the foliar spray of date palm for improvement of mineral composition in the date palm plant. Starting in 2007, the effect of urea (N 46%) has been observed when sprayed on the plant in different concentrations (0.5% and 1%) and in different stages of the plant (Khalal and Kimmri stages) (Abbas et al. 2007; Khayyat et al. 2007; Shareef 2011a, b). Positive traits of plant growth on the same included an increase in the pulp weight, nitrogen content, fruit length and diameter, dry matter, percentage of fruit ripening, and sugar content while decreasing dropping down of fruits from the plant. Similar results were obtained using 300–600 ppm zinc sulfate in terms of yield and quantity of fruits. Total soluble solids calculated were higher using 1500–2500 ppm boric acid along with the increased concentration of boron content in the fruits. Moreover, other nutrients like 1000–2000 ppm of calcium chelate were responsible for the increase in plant height, girth, and the number of new leaves (Jasim et al. 2016). An overall increase in fruit length and diameter, along with the increased percentage of sucrose and total dissolved solutes, was obtained by application of 125 and 250 mg/L phosphorus (Fasal et al. 2014).

It is essential that extra importance is laid on the status of nutrition of any crop and biofortification plays an added role in this process. A constant effort to elevate nutritional content in date palms has been initiated by several scientists across the globe, and one of the many noteworthy discoveries includes the introduction of (0.01–0.02%) sodium selenite for improvement in growth, bunch weight, and quality of fruits in the plant (El-Kareem et al. 2014). Nitrogen, phosphorus, and potassium (NPK) are mixed in various ratios to fasten plant growth and for other beneficial properties. In the cultivation of date palms too, NPK was prepared in the ratio of 20:20:20 and in 2.5% concentration which provided beneficial effects like increase in total yield; soluble solids; sugars; percentage of fruit ripening; chlorophyll content; soluble carbohydrates; potassium, phosphorus, and nitrogen content; chlorophyll a; fruit yield; and dry weight and decrease in the number of fruit drop (Shareef 2011a, b; Attaha and Al-Mubark 2014; Altememe et al. 2017; Jubeir and Ahmed 2019). Thus, usage of NPK is beneficial to a high commercial extent but the side effect of chemical pollution and chemical fertilizers always persists in this case, which may prove harmful for human consumption. It is well-known that the role of

potassium in plants ranges from the transport of minerals, water, and carbohydrates and the production of protein, starch, and adenosine triphosphate to the improvement of drought resistance and others. In addition, various salts of potassium have been tested on the growth of date palms in various concentrations such as potassium sulfate, silicate, nitrate, and others. Potassium nitrate when used in the concentration of 1500 ppm exhibit a notable increase in the number of new leaves and clusters along with the increase in the content of ascorbic acid, abscisic acid, indole acetic acid, carbohydrates, gibberellin, and zeatin (Shareef 2019). When potassium silicate was used in concentrations of 0.01% and 0.02%, an increase in bunch weight and percentages of specific minerals in leaves, yield, and leaf area was noted (El-Kareem et al. 2014). Among pigments, there was higher chlorophyll a, carotenoid, and total chlorophyll when the same was applied. Similarly, the utility of iron compounds in the production of chlorophyll and other chloroplast maintenance activities in plants cannot be ignored. Iron sulfate itself when used in 250 ppm concentration can lead to a good increase in yield along with a decrease in the number of fruit dropping (Abbas et al. 2007). Again, when the same compound is used in 20–40 ppm, it results in an increase in weight and volume, diameter, length of fruits and flesh, carbohydrates, total soluble solids, reducing sugar, and dry weight (Abass et al. 2012). Apart from iron sulfate, Faisal et al. (2017, 2018) have reported that the usage of chelated iron compounds in date plants has led to an increase in the overall total soluble solids, reducing sugar, weight, and diameter of fruits.

The list of agronomic biofortification methods continues to include the usage of 8–16 cm^3/L seaweed as a biostimulant in date palm plants (Attaha and Al-Mubark 2014). It produced an increase in a long list of parameters such as soluble carbohydrates, potassium, phosphorus, as well as chlorophyll content. When the same biostimulant was used in the concentration of 4–8 ml/L, the effects observed in the fruits included elevated levels of chlorophyll, K, Mn, ascorbic acids, organic solutes, indole-3-acetic acid, zeatin, K^+/Na^+ ratio, and other ions as compared to that without seaweed (Taha and Abood 2018; Shareef et al. 2020). Shareef et al. (2020) also reported that 4 g/L *Saccharomyces cerevisiae* can bring about almost the same effects with an additional benefit of increase in plant height, new leaves, offshoot girth, and leaf area. Apart from this, the gelatinous milky white secretion produced by worker bees, also called "royal jelly," is responsible for the production of high nutrient content like N, P, K, and Mg along with good quality fruits and increased growth (Al-Wasfy 2013). Refaai (2014) reported that using 0.5–2% wheat seed sprouts increase growth characteristics, plant pigments, total yield, nutrients, bunch weight, and carbohydrates along with improvement in characteristics of fruits – both physical and chemical. Macro and micronutrients together are essential for the plants to grow healthy and the higher the quantity of nutrients available to the plants, the better growth it exhibits. Through biofortification, the level of mineral uptake is maximized in plants and one such effort was put forward by combining many micronutrients like Fe, Zn, Cu, Mn, B, and Mo and chelating with EDTA. Here, boron and molybdenum are used in completely soluble form, under the name of "Fetrilon Combi," and is commercially available as an agronomic biostimulant in crops (Altememe and Mahdi 2015). Using the product shows a remarkable increase

in the invertase activity, ripening percentage, and ionic concentration of sucrose and water content in fruits. Another commercialized biostimulant available in the name of "Drin" (2–4 ml/L) helps date plants to synthesize free amino acids and readily uptake nitrogen content from the soil. Catalase activity is increased along with total chlorophyll content as this liquid organic solvent contains a high proportion of amino acids and organic nitrogen that activates the biochemical pathways in crops. Like many other biostimulants, addition of 4–8 ml/L humic acid is not only seen to increase the amount of N, P, and K but also the overall content of chlorophyll a (Altememe et al. 2017). Obtained from animal bones, 5–10 ml/L sugar chitosan is a good source of increasing K, Mn, amino acids, and dry weight of fruits in date palms (Taha and Abood 2018). Recent investigations also reveal the usage of 1.5 g/L commercialized Biocont-T in date palms shows benefits like increased bunch weight, yield, and percentage of fruits and 800 or 1600 mg/L potassium fertilizer by Alqawafel shows notable increases in fructose, glucose content, and fruit size with reduced sugar (Ressan and Al-Tememi 2019; Al-Hajaj et al. 2020).

Date palm is an essential source of food in areas where there is water scarcity and thus demands a good number of postharvest techniques for their growth and biofortification. Among a wide range of technologies available, storing at −18 °C and application of apple vinegar treatment have proven to be the most useful in effect (Kahramanoglu and Usanmaz 2019). Next effective in line are the treatments of grape vinegar followed by commercialized Ethephon; the former is credited with an increase in total soluble solids in the fruits, while the latter slightly affects the ripening of fruits due to reaction with tannins.

2.4.2 Postharvest Techniques

After harvesting a crop, it is essential to protect and conserve the agricultural produce with minimum handling and effective management. This is mainly done using postharvest management strategies and employs many advanced scientific and technical processes. A good harvesting procedure and management can lead up to 100 kg of dates annually where the average economic life of a date garden can be around 150 years. While harvesting, it has been found that usage of blue and black polythene bags has increased the percentage of fruit ripening (Awad 2007). Similarly, it is advisable to implement the bagging process to protect the plants from birds, sunburn, and heavy rain (Zaid and de Wet 2002b). The grower's experience plays a big role in deciding the time of harvest of the fruits depending on their texture and appearance which are connected to sugar content and moisture. If decided correctly, the timing of harvesting reduces chances of skin cracking, attacks by insects and pests, or extreme dehydration. In fact, when selectively harvested early, it also leads to higher prices in the market and avoids disadvantages of harvesting late such as adverse weather conditions, attacks by pests, yielding good quality dates at matured stages, etc. However, in case the dates fall on the ground, they are not supposed to be considered fit for human consumption as there are higher chances of microbial contamination and soil particles entering the date fruit (Kader and Hussein 2009).

Artificial Ripening

It is very important that the fruits harvested are taken for appropriate ripening if they are picked immature for any purpose. Ripening can be carried out either indoors or outdoors, where the former requires a room with adequate air-conditioning to keep temperatures at 45–46 °C and 70% relative humidity, for dates with thick flesh to ripen in about 2–4 days (Hyde 1948). However, the latter is a cheaper technique and comes with many disadvantages such as adverse conditions or external attack. Apart from this, the quality of ripening further increases when fruits are treated with chemicals like 1% ethanol and 2% acetic acid +1% NaCl or acetaldehyde (Asif and Al-Taher 1983). It has also been observed that freezing at −35 to −50 °C is beneficial for the plants as it prevents damage to cell membranes (Kader and Hussein 2009).

Hydration

The step necessary to soften the few hard fruits refers to the hydration process which includes dipping them in cold or hot water for around 4 to 8 h at 60–65 °C (Kader and Yahia 2011). The relative humidity maintained in the same is around 100% to convert the dried dates into glossy ones. Maintaining a constant rate of temperature and relative humidity is essential through forced air circulation, which is also the method of controlling the growth of microorganisms. The methods of treatments vary from one country to another, depending on the climatic conditions. Chemicals like alkaline ammonium sulfate help in improving quality of acidic but hydrated dates (Yahia et al. 2014). The additional step of hydration is necessary only when the dates are unripe and not overdried. Pasteurization is an exclusive process in the postharvest treatment of date palms which is executed at 72–80 °C for 2.5 min (Martínez Vega et al. 2014). It was also reported that maintenance of almost 15–70% of relative humidity is best, when keeping the fruit mass loss and sweetness index in mind along with the coloration of dates becoming darker when skin temperature reached around 66 °C (Casagrande et al. 2021). Thus, hydration is essential only when the moisture content of the fruits is low and can be continued using the same solutions as applied during the harvest procedure (Dole John and Faust James 2021).

Initial Transportation and Sorting

Large boxes of wood, plastic, or cardboard capable of carrying 200–450 kg fruits are essential to prevent any postharvest losses while transferring the fruits to the packing station. The process is better when faster and speedy transportation can be done as that prevents microbial infection as fruits are more susceptible to pests during the postharvest period (Abd Elwahab et al. 2019). The transportation process is ideally carried out at cooler temperatures (0–2 °C) to prevent spoilage and atmospheric heat damage to plants. Specifically, forced air cooling proves to be more efficient than hydrocooling, which is the removal of excess moisture and its disinfection before packing and the usage of perforated plastic lining adds to the technicality of the process (Kader and Hussein 2009). Further, dates are fumigated in closed chambers and transferred to shakers for primary washing followed by the removal of excess water from the fruits. Then they are sorted manually through visual inspection

according to whether they are ripe or unripe and are discarded if they are damaged due to transportation, microbial infections, heat, or other external factors, so that only the fresh lot of fruits is kept (Sarraf et al. 2021). This is one of the most time-consuming stages of all postharvest procedures because sorting and grading depend on either manual means or through a mechanized visual system where color values and sugar range play an important role in differentiating various types of dates.

It is essential to remove those dates which are either unfertilized or disfigured or contain some blemishes or foreign particles. This can be executed manually or using a machine according to physiological parameters like color, shape, size, and others. Immature, overripe, diseased, parthenocarpic, and partially or completely damaged dates are ideally discarded in this stage and only fresh ones are sorted according to their quality. Date fruit colors range from yellow to brown, and consistency of flesh ranging from soft to dry; these are the two most essential characteristics at this point (Sarraf et al. 2021). The accuracy of grading was improved using automated systems like date quality analysis by near-infrared imaging of two-dimensional view (Lee et al. 2008). More advanced techniques are also being developed for the differentiation of a variety of species since the time of maturation of all fruits is not the same, which will hasten the entire process of sorting and transportation to the packing house.

Advanced Automation and Robotics in Fruit Handling

Keeping in mind the disadvantages of water and soil unavailability, it is always advisable to make the optimum use of advancements in technology and automation, which not only increases the rate of accuracy but also speeds up the entire procedure in a limited area. Cull fruits can be discarded or storage of fruits for a longer period can be controlled by automated systems and algorithms fixed in advance. This highlights the necessity of attaching technology to agriculture, forming agri-tech and taking the help of GPS, the Internet, robots, sensors, and a many other electronic devices to execute the agricultural processes in a more organized, smart, and accurate manner (Rose and Chilvers 2018). The remarkable possibilities of the Internet going digital are well-known in the twenty-first century and in the whole world; thus, it is time to transform conventional procedures into smarter and faster techniques. Near-infrared spectroscopy has long been helping in the analysis of fruit colors and textures (De Jager and Roelofs 1996). Not only this, but the advancement has also reached to the level that packing lines containing several sensors and cameras attached to visible and infrared probes help in analyzing the different qualities and parameters of fruits (Walsh et al. 2020). A track record can be maintained for easy accessibility (Njoroge et al. 2002) of the intricate details of the fruits like color, shape, size, grower name, and others by consumers to guarantee the safety and security of the fruits. Furthermore, computer models are used for correct decision-making and 80% accurate grading of fruits according to RGB images produced (Al Ohali 2011). There can be no end to the innovation and technology when it comes to applying robotics and artificial programming for both consumer's and manufacturer's ease of transaction.

Nanotechnology Utilization in Packaging

The application of nanotechnology has become a well-researched topic in almost every sphere of life, from physics to the biomedical world. For the packaging of date fruits, nanotechnology can play a major role in packaging, coatings, and films for date fruits. There are various types of nanoparticles – metallic, nonmetallic, or synthesized – working as antifungal agents in the preservation of fruits. There is a provision for creation of edible packages also using chitosan-based nanoparticles, which helps in the reduction of postharvest decay in fruits (Chaudhary et al. 2020; Kumar et al. 2020). Additionally, respiration and transpiration rates can be controlled in date fruits by using nanotechnology materials with smart packaging and innovative technologies for increasing the shelf life of fruits when stored for long periods (Acharya and Pal 2020; Alfei et al. 2020). Thus, overall application of nanotechnology using nano-fertilizers can also aid in the process of postharvest management of date palms to minimize the losses incurred by it.

Adding Surface Coatings

It is essential to identify the presence of any metal particles or xenobiotic compounds present on the fruit that might have come from the environment. This is followed by the improvement in the appearance of dates by introducing polished surface coatings and reducing stickiness. In addition, a solution of 5–6% soluble starch can be used as a dipping agent, 3% methylcellulose, 6% vegetable oil, 2% mixture of butylated hydroxyanisole and hydroxytoluene, 90% water, and finally a wetting agent. Ethanol vapors or immersing the dates in water for 10 h has hastened the process of ripening as compared to the controls, and neither had any negative effect on the quality of fruits (Awad 2007). For the surface coating preparation, solutions of date syrup, vegetable oil, corn syrup, and sorbitol or glucose syrup can also be used.

Cooling and Packaging

As mentioned, hydrocooling, which affects the rate of microbiological and biochemical changes in temperature and maintaining date palms at an average temperature of 0–10 °C, is always desired and recommended for further long-term storage under 65–75% relative humidity. Once removal of excess water and disinfection is done, dates can be cooled to 0 °C in 10–20 min, based on their initial temperature. Hydrocooling delays the rate of spoilage caused by microbes and increases the shelf life of food products (Djihad et al. 2021). Maintaining the temperature is important as it affects physiological parameters like sugar crystallization and that is affected by the moisture content in the fruit, as water above 20% affects the fruits adversely (Glasner et al. 1999).

Before commercialization, it is essential that the date palms are packed in suitable containers or packages so that they are protected from any further microbial attack or damage due to mishandling and transportation prior to consumption. Only those date palms that are packed and available in intact airtight packets are accepted for further research and analysis in the marketing sector (Kabir et al. 2021). Consumer packages and large cartons including transparent film bags and plastics include the containers for storing date palms apart from other bottles with metal caps and bottoms. Small

packages suited to carry 50–60 g date palms are also used along with cardboard boxes weighing 5 kg, in countries like the USA, Jordan, Saudi Arabia, and others. It is very important to mention particulars about the specifications on the labels such as date of packaging, grower industrial details, nutritional benefits, date of expiry, country of origin, and other details so that consumers can get a glimpse of the overall status of the fruit.

2.5 Requirement of Genetic Biofortification

Fortification of any consumable product refers to deliberately making it enriched and more nutritious than normal by means of either the addition of micronutrients or bringing about modification genetically. It is done either through conventional techniques which are slower and vulnerable to bacterial invasions or through modern biotechnological approaches that are quicker, more reliable, and easier to obtain. Such an approach toward the biofortification of any consumable is essential when judged from the perspective of enriching the nutritional status of any country's population. Processes of agronomic practices, food processing, and others fall under conventional fortification techniques, but for genetic modification of plant germplasm, selective breeding comes under modern biofortification tools and techniques. It is aimed toward the enrichment of a nutritional density of a plant, increasing accessibility to rural areas without the slightest drop in the overall yield of the crop. Especially after the COVID-19 pandemic, it has become essential to support the underdeveloped and developing countries in terms of food security and uplift their standards of living in terms of nutrition and health to prevent loss of life from malnutrition. Moreover, institutions like ICRISAT and CIMMYT in association with Harvest Plus have taken initiatives to scale up biofortification procedures to improve the nutrition and public health by providing and promoting higher iron content in pearl millets and zinc content in wheat available to the general population (https://www.harvestplus.org/where-we-work/india). India ranks tenth out of 117 countries and third out of 128 countries, respectively, as suitable place to grow iron pearl millet and zinc wheat. Increasing food security and the quality of life of people in developing countries and decreasing nutrition-related disorders, mortality, and morbidity are the main goals of biofortified crop production.

When crops are genetically engineered to produce high micronutrient-enriched products, they are referred to as genetically fortified. One of the most common examples of genetically biofortified crops is the Golden Rice that was modified to produce and accumulate vitamin-A enriched genes (β-carotene), giving it a golden colored appearance. Rice being the staple food of a good portion of the entire world population, if targeted to be genetically biofortified, can automatically benefit a lot of people who might not get the required nutrients due to the reduction in zinc and iron, especially after milling. So, one of the most promising solutions brought very recently toward the mitigation of this problem is the production of genetically biofortified rice along with iron and zinc, and in addition special features were kept in mind, for the decrease in arsenic contamination in rice grains (Viana et al.

2021). Not only rice, but there are also reports of wheat biofortification as wheat is also consumed by another major part of the population. Identification of Fe/Zn QTLs is an important step of biofortification, and a versatile domain is promised by the usage of biofortification tools and techniques in genetic engineering and molecular biotechnology field (Tong et al. 2020). Genomic selection has been considered as a good process for biofortification of Fe/Zn QTLs, but incorporation of the QTLs into the models as a fixed effect is questionable. The answer to why crops need to be genetically biofortified can be alluded to targeting the rural population which has restricted accessibility to nutritious foods and minerals or those who have almost no opportunity of genetic interventions like biofortification. Another reason is the requirement to meet a specific quantity of calories needed by the body to sustainably live.

As of recent reports, Egypt has recorded the world's highest date palm consumption of about 1.6 million mt, followed by Oman, United Arab Emirates, and Algeria (https://www.mordorintelligence.com/industry-reports/date-market). Africa and Middle East alone hold 70% of the entire consumption of date palm fruits worldwide and the trend increases daily. So, with the increase in date fruit consumption, the necessity to biofortify the fruit increases and achieving the target of uplifting lives in rural areas with micronutrient-enriched date fruits automatically becomes easier. The proteins like cell number regulator (CNR) play an important role in the genetic biofortification of *Triticum aestivum* by TaCNR5 expression in shoots of wheat analyzed under Cd, Mn, or Zn metal treatments (Qiao et al. 2019). Activity of these heavy metals as micronutrients in the biofortification of cereals, moderated using the expression levels of TaCNR5, can open a new avenue of research and development in the nutrition elevation. As of 2021, not enough research has been executed for the genetic biofortification of date palm, but the techniques are being studied and more possibilities are being discovered by plant biotechnologists. In their first book, Al-Khayri et al. (2021) provided comprehensive information on various proteomics and metabolomics technologies applicable for the genomic improvement of date palm fruits. The genomic stability of in vitro date palm plants and diversity of genomes with DNA barcoding processes along with genomic approaches for resistance to all stress factors have been reported already. At this juncture, this chapter proposes the urgency and essence to identify the possible genetic biofortification procedures applicable for date palm because that itself serves as one of the major mechanisms for nutrient level improvement in the world population.

The requirement of genetic biofortification alludes to the necessity of nutritional security in a country, with respect to enhancement of the micronutrient status in the soil. Transgenic biofortification is a route that can answer the same question and curtail the issue of micronutrient deficiency in soil. At this juncture, microbes with plant growth-promoting benefits are beneficial to improve the nutritional quality of the soil. Reports of using *Arthrobacter* sp. (DS-179) and *Arthrobacter sulfonivorans* (DS-68) for the nutritional improvement of zinc and iron in the soil have been studied (Singh et al. 2017). Biofortification, depending on the insert type, can be classified into conventional and transgenic biofortification. The former type involves the introduction of artificial additives along with it and is thus largely limited to the

developed countries only, due to the necessity of high infrastructure and machinery (Díaz-Gómez et al. 2017). There are various disadvantages connected to the introduction of added minerals to the food crop like taste, etc., but the brighter side includes a higher percentage of necessity, when focusing on the nutritional requirements among the masses. The importance and necessity of genetic biofortification, though well-known to scientists and researchers, has not much been implemented in the field of date genomics. This void needs to be filled through advanced research as soon as possible and that demands our attention to elevate the standards and processes that are linked with enhancement of mineral constitution in the plant at study.

3 Breeding Using Molecular Markers for Genetic Improvement

Molecular breeding via molecular biology and marker-assisted DNA analysis by considering phenotypic character and diversity is an effective strategy to analyze genetic polymorphism in breeding programs. Molecular breeding includes marker-assisted selection (MAS), quantitative trait loci (QTL), genetic engineering, etc. Molecular markers with support of phenotypic data are involved in species and cultivar identification, diversity and phylogenetic relationships, and many breeding activities.

3.1 Diversity Analysis

3.1.1 Phenotypic Diversity

Phenotypic diversity is seemingly a date palm diversity indicator. It is also a basis for selection, conservation, and improvement for sustainable utilization. There are 19 known or reported *Phoenix* species; however, 12 species (Zabar and Borowy 2012; Abul-Soad et al. 2017) are often considered as valid as there is still confusion regarding the exact number of *Phoenix* species. *P. dactylifera*, *P. acaulis*, *P. canariensis*, *P. paludosa*, *P. reclinata*, *P. rupicola*, and *P. sylvestris* are the most widely accepted species. *P. dactylifera* is the most commercially cultivated *Phoenix* species. Pinnate leaves and conduplicate leaflets with acute apex are the distinguishing characteristics of this genus (Uhl and Dransfield 1987). *P. dactylifera*, *P. atlantica*, *P. canariensis*, *P. theophrasti*, and *P. sylvestris* are closely related in overall appearance. However, they have been reported as separate species by molecular marker studies (Henderson et al. 2006).

P. dactylifera is the original Middle Eastern wild representative and is distributed in different parts of South Asia and Africa. The domesticated date palm grows more than 30 m in height. It has a clustering trunk, but in cultivated varieties, it appears as a single trunk due to the removal of offshoots. The fruits are the largest among the *Phoenix* spp. reaching to a maximum of 100 × 40 mm in size. The other wild relatives like *P. theophrasti* have a restricted distributional range especially in coastal

areas of Southwest Turkey, *P. atlantica* growing near the Atlantic shores of North Africa, which represents the feral populations of *P. dactylifera* (Rivera et al. 2008), and *P. iberica* with glaucous leaves, stout stem, and small fruit with thin flesh growing in the Mediterranean coast of Spain which is considered as an intermediate between *P. theophrasti* and *P. sylvestris*. *P. chevalierii* has greener leaves compared to *P. dactylifera* and is an Iberian-Moroccan group of cultivars grown in southeastern Spain (Rivera et al. 2008). *P. reclinata* has a bushy habit and seen in the southern parts of Arabia and Sahara Desert (Tengberg 2012), and *P. sylvestris*, the rain palm, is a nonsuckering 20 m tall tree adapted to tropical climates especially the Indus Valley (Zohary and Hopf 2000). *P. canariensis*, usually cultivated as an ornamental, has a 20 m high stout trunk which is adapted to moderate climate and is endemic to the Canary Islands (González-Pérez et al. 2004). However, *P. rupicola*, known as the cliff palm is about 7 m in height and has a thin trunk and is native to the northern parts of India. Similar in morphology is the *P. pusilla* palm but with a shorter trunk and is found in the southern parts of India and Sri Lanka. The Senegal date palm *P. reclinata* with 10 m thin clustering trunk is used as an ornamental and is native to tropical Africa. Ornamental palms *P. paludosa* (mangrove date palm) and *P. reclinata* have similar appearance and are found in the swampy regions of southeastern Asia. *P. roebelenii* (pygmy date palm), grown as an ornamental, has a single trunk and is native to Southeast Asia; *P. acaulis*, *with* short clumping stems, is found in northern India and Burma; and *P. loureiroi*, often confused as *P. acaulis*, has short stems and is found in northern India and southern China (Barrow 1998; Zohary and Hopf 2000).

Thousands of cultivars have been reported from all over the world. These have been developed by various selection methods to improve crop yield and quality and to tolerate environmental stress, etc. Morphological description of characters like form of the tree, fruit (shape, size, weight, color, texture), fruit stalk, and leaf, especially leaflets and spines, is an important parameter for identification of cultivars (El-Houmaizi et al. 2002; Al-Yahyai and Al-Khanjari 2008; Ould Mohamed Salem et al. 2008; Elshibli and Korpelainen 2009; Markhand et al. 2010). Jaradat and Zaid (2004) have reported up to 5000 date palm cultivars. However, based on botanical descriptions, 1000 cultivars in Algeria, 450 in Saudi Arabia, 400 in Iran, 400 in Iraq, 250 in Tunisia, 244 or 453 in Morocco, 95 in Libya, 400 in Sudan, 250 in Oman, 321 in Yemen, 52 in Egypt, and 300 in Khairpur, Pakistan (Benkhlifa 1999; Bashah 1996; Hajian and Hamidi-Esfahani 2015; Zabar and Borowy 2012; Zaid and de wet 2002a, b; Sedra 2015; Battaglia et al. 2015; Osman 1984; Elshibli 2009; Al-Yahyai and Al-Khanjari 2008; Al-Yahyai and Khan 2015; Ba-Angood 2015; Rabei et al. 2012; Mahar 2007; Markhand et al. 2010; Abul-Soad et al. 2015;) along with other cultivars in different growing regions have been reported.

Chemical characters of fruits and ripening stages of fruits are some of the apparent factors for the identification of cultivars (Elshibli and Korpelainen 2009, 2010). Numerous documented reports are available for date palm cultivars using vegetative, flowering, and fruit characters. In Saudi Arabia, 17 date palm cultivars were evaluated for vegetative parameters, flowering and yield characters, and fruit attributes (Al-Doss et al. 2001). Among five Sudanese date palm cultivars,

physicochemical differences with respect to weight of fruit, and seed, thickness of flesh, fruit length, and phytochemical constituents were reported (Sulieman et al. 2012). Prominent fruit characters were reported among 85 cultivars from Pakistan (Markhand et al. 2010). Twenty-six date palm cultivars have been studied to identify descriptors (leaf, pinnae, and spine) for early-stage date palm characterization other than fruit characters (Elhoumaizi et al. 2002). Fruits of date palm have been classified based on texture as soft, semidry, and dry types (Barreveld 1993). Ecological distribution of soft and dry dates is reported only from Sudan. However, they also exist in different growing countries (Barreveld 1993; Zaid and de Wet 2002a). Although there is huge diversity in the cultivar of date palm, most information is unpublished or unavailable due to ownership of cultivars.

3.1.2 Genetic Diversity Using DNA Markers

Phenotypic characterization has been utilized extensively for the identification and selection of various cultivars. However, it is difficult to identify the date cultivars without observing the fruiting stage. Molecular biology techniques and molecular marker-based selection have helped breeders to accurately select specific cultivars. The extent and distribution of genetic diversity is an important measure for genetic conservation of existing germplasm (Jubrael et al. 2005). The availability of nuclear and chloroplast genome sequence of date palm is an added advantage for assessment of genetic diversity (Yang et al. 2010; Al-Dous et al. 2011; Soumaya et al. 2014). In addition, molecular markers can be effectively used for genetic diversity assessment and constructing genetic maps for identification, genetic improvement, and conservation of true-to-type elite material (Eissa et al. 2009; Khierallah et al. 2011).

Random Amplified Polymorphic DNA (RAPD)

RAPD marker-based PCR techniques have been extensively used for identification of date palm cultivars as they are cost-effective and do not require blotting or radioactive materials (Mirbahar et al. 2014; Emoghene et al. 2015). Numerous studies have reported genetic diversity in date palm cultivars using RAPD markers in Saudi Arabia, Tunisia, Algeria, Iraq, Egypt, Bahrain, Syria, Nigeria, Morocco, India, and Pakistan (Al-Khalifah and Askari 2003; Al-Moshileh et al. 2004; Abdulla and Gamal 2010; Munshi and Osman 2010; Benkhalifa 1999; Jubrael 2001; Al-Khateeb and Jubrael 2006; Khierallah et al. 2014; Trifi et al. 2000; El-Tarras et al. 2002; Adawy et al. 2004; Soliman et al. 2006; Younis et al. 2008; Moghaieb et al. 2010; Pathak and Hamzah 2008; Haider et al. 2012; Sedra et al. 1998; Sedra 2013; Emoghene et al. 2015; Mirbahar et al. 2014; Toor et al. 2005; Singh et al. 2006; Rani et al. 2007). However, due to lack of dominance and detection of heterozygosity, these markers are relatively less used than other molecular markers.

Amplified Fragment Length Polymorphism (AFLP)

Large intervarietal polymorphism in date palm cultivars has been reported by various workers using AFLP markers (Mueller and Wolfenbarger 1999; Elhoumaizi et al. 2006; Khierallah 2007). These markers have also been effectively employed for making high-density linkage maps. There are several reports showing the detection

of polymorphism in date palm cultivars. In Spain, Diaz et al. (2003) reported diversity and polymorphism of date palm cultivars using AFLP markers. Using AFLP markers, El-Khishin et al. (2003) characterized five date palm cultivars from Egypt. The genetic diversity of Egyptian date palm from eight locations was studied. Twenty-eight accessions and a few accessions from California were classified and found to represent the major date palm germplasm of North Africa (El-Assar et al. 2005). Genetic diversity and phylogenetic relationships among various Iraqi cultivars have been reported (Jubrael et al. 2005; Khierallah et al. 2011). AFLP markers have been also employed for assessing genetic variations that are seen in offshoots of tissue-cultured date palms (Saker et al. 2006).

Restriction Fragment Length Polymorphism (RFLP)

RFLP markers are locus-specific markers which can be used as important tools for identification between closely related species. Being codominant they show strong molecular differentiation. Date palm cultivars have been differentiated using these RFLP markers. In Tunisian date palms, 43 accessions have been utilized for identification of genotype and genotypic polymorphisms (Sakka et al. 2003). Shoot tips used to initiate tissue culture of five elite date palm cultivars were utilized to determine the polymorphism using RFLP markers (Cornicquel and Mercier 1994). Cornicquel and Mercier (1997) generated cultivar-specific hybridization using RFLP in four date palm cultivars. PCR-RFLP-based markers were utilized for sex determination in date palm (Al-Mahmoud et al. 2012).

Intersimple Sequence Repeats (ISSR)

Microsatellite-based primers are utilized to amplify intersimple sequence DNA repeats and have been also utilized for studying inter- and intraspecific genetic variations in date palm. The genetic diversity of date palm in different growing regions has been assessed by various workers in Egypt, Iraq, Iran, India, Morocco, Pakistan, Tunisia, Egypt, Saudi Arabia, Algeria, and Ethiopia (Younis et al. 2008; Khierallah et al. 2014; Sharifi et al. 2018; Srivastav et al. 2013; Bodian et al. 2012a; Mirbahar et al. 2013; Ahmad et al. 2020; Zehdi et al. 2002, 2004a, b, 2005; Karim et al. 2010; Zehdi-Azouzi et al. 2011; Hamza et al. 2012; Younis et al. 2008; Kumar et al. 2010; Munshi and Osman 2010; Sabir et al. 2014a; Boudeffeur et al. 2021; Takele et al. 2021).

Microsatellites or Simple Sequence Repeats (SSR)

SSR markers are codominant, species specific, and highly polymorphic, making them suitable markers for cultivar identification, genetic diversity assessment, and construction of linkage maps and gene-based maps in date palm. Akkak et al. (2003) have isolated simple sequence repeats from genomic library and detected high polymorphism in the analyzed samples. Researchers have reported genetic diversity and polymorphism in Tunisia (Zehdi et al. 2004a, b, 2006), Qatar (Elmeer and Mattat 2012), Iraq (Khierallah et al. 2011), Libya (Racchi et al. 2014), Saudi Arabia (Yusuf et al. 2015; Al-Faifi et al. 2017), Qatar (Ahmed and Al-Qaradawi 2009; Elmeer and Mattat 2015), Mauritania (Bodian et al. 2012b), Pakistan (Naeem et al. 2018), and

Sudan (Elsafy et al. 2016). SSR markers have been developed for specific identification of gender in date palm for breeding (Maryam et al. 2016).

Expressed Sequence Tags (EST)

28,889 EST sequences were analyzed from date palm genome database to develop gene-based markers, i.e., EST-SSRs (Zhao et al. 2012). One third of the primers designed detected polymorphism to differentiate the date palm cultivars used. In date palm, large-scale collection and annotation of gene models were also developed (Zhang et al. 2012). ESTs were also generated for understanding the high performance and quality of commercial cultivar (Al-Faifi et al. 2017).

Single Nucleotide Polymorphisms (SNPs)

High-density genetic maps are created using potential markers like SNPs. They are now widely used to study date palm genome sequences. The first draft genome of palm cultivar was assembled using EST markers (Al-Dous et al. 2011). SNP analysis by sequencing the mitochondrial DNA sequence of Saudi Arabian date palm cultivars revealed close phylogenetic relationships among the studied cultivars (Sabir et al. 2014b). In a study of date palm cultivars from different countries using generated SNPs, the date palms were segregated based on genetic background into two main regions, i.e., North Africa and Arabian Gulf (Mathew et al. 2015). Based on whole genome sequence of date palm, a catalog of about seven million SNPs from 62 cultivars was developed (Hazzouri et al. 2015).

3.2 Sex Determination

In dioecious trees, identification of male and female sex is a complex process, as it is not possible to identify the plants in the seedling or early growth stage. Date palm being dioecious takes 5–7 years before it starts flowering and fruiting, which is a limiting factor with respect to traditional breeding programs. It not only causes economic loss to farmers but also makes it difficult for the breeders to select the superior lines from the existing elite cultivars. The genetics of sex determination in date palm is fully understood, but attempts have been made at morphological level, biochemical level, and molecular level to determine the sex of the plant at the early development stages (Awan et al. 2017).

3.2.1 Morphological Markers

Morphological features of the plant are usually taken into consideration to determine the sex of the plant. However, such characters can be applicable at maturity. Being laborious and the effect of environmental factors make it difficult to study. Identification of genotype of date palm is based on female tree morphology and characters of fruits. Date palm cultivars from Saudi Arabia have been identified based on morphological characters (Al-Khalifah et al. 2012). Similarly, male plants from elite cultivars have been identified based on morphological markers (Soliman et al. 2013). There are

numerous reports stating that the leaf and leaflet characters can be utilized for identification (Ahmed et al. 2011; Hammadi et al. 2009; Haider et al. 2015).

3.2.2 Biochemical Markers

Peroxidase serves as an important marker for the identification of male and female plants, as the diversity of peroxidase varies between male and female inflorescences. Researchers have reported lower peroxidase and acid phosphatase activities in male than in female date palm plants (Qacif et al. 2007; Bekheet et al. 2008). Moreover, glutamate oxaloacetate activity is higher in female than in male plants. Polypeptides specific to male plants which were isolated from the leaves were also reported. These also serve as a gender biomarker in date palm (Sonia et al. 2013). Chemical composition of leaves has also been a distinguishing parameter, where female leaves have higher concentration of pigments, phenolic acids, amino acids, and sugars, whereas male leaves have higher concentrations of proline and ash (El-Yazal 2008). Higher levels of sugar have been reported in male trees compared to female trees (Rao et al. 2009). Similarly, phenolic acid content is higher in male than in female plant saps (Makhlouf-Gafsi et al. 2016).

3.2.3 Molecular Markers

Due to the limitations of the morphological and biochemical markers, molecular markers based on RFLP, RAPD, AFLP, and SSR markers were developed for the identification of gender in date palm. Numerous attempts to identify the male and female gender in date palm using RFLP (Al-Mahmoud et al. 2012), RAPD (Ben-Abdallah et al. 2000; Soliman et al. 2003; Ahmed et al. 2006; Al-Khalifah et al. 2006; Bekheet et al. 2008), ISSR (Younis et al. 2008; Al-Ameri et al. 2016a), AFLP (Atia et al. (2017), SCAR (Al-Qurainy et al. 2018; Dhawan et al. 2013; Al-Ameri et al. 2016b), and SSR markers (Elmeer and Mattat 2012; Maryam et al. 2016) have been made. A putative sex chromosome was identified using the constructed date palm genetic map (Mathew et al. 2014). Date-SRY gene and GWAS mapping of sex determination locus are some of the important tools for sex determination in postgenomic era (Hazzouri et al. 2019; Mohei et al. 2019).

3.3 Genomics

Sequencing and analysis of the genome have important applications in determining genetic diversity, systems biology, evolutionary process, etc. The sequenced genome of date palm has been used for developing genetic map (Al-Dous et al. 2011; Al-Mssallem et al. 2013; Mathew et al. 2014; Sabir et al. 2014b; Hazzouri et al. 2015). Sequencing technologies like pyrosequencing, ligation-based sequencing, etc. have been utilized for sequencing the nuclear genome (Al-Mssallem et al. 2013), chloroplast genome (Yang et al. 2010), and mitochondrial and plastid genomes (Sabir et al. 2014b). In addition, to understand the genomics of date palm, various projects have been undertaken for sequencing the whole genome like the Date Palm Genome Project (DPGP) by King Abdulaziz City for Science and Technology (KACST), Riyadh, Saudi Arabia, in association with Beijing

Institute of Genomics in 2008. Weill Cornell Medical College, Qatar, using next-generation sequencing technology, sequenced the whole genome of date palm (https://qatar-weill.cornell.edu/research-labs-and-programs/date-palm-research-program/date-palm-genome-data). The genome contains 500 Mbp.

3.4 Computational Analysis

Gene regulation has been highly influenced by micro-RNAs (mi-RNAs) during the development of the plant. The 276 novel date palm-specific mi-RNAs involved in regulating genes for fruit development and ripening have been characterized by using high-throughput sequencing and bioinformatics predictions (Xin et al. 2015). Similarly, 153 homologs of conserved miRNAs, 89 miRNA variants, and 180 putative novel miRNAs in date palms that can play an important role in salt tolerance were identified (Yaish et al. 2015).

3.5 Genetic Manipulation of Date Palm

Date palm has been genetically modified to incorporate new genes to improve the commercial value of the crop. Transformation via *Agrobacterium-mediated* and direct gene transfer using particle bombardment has been utilized for the transfer of specific genes of interest. *Agrobacterium*-mediated transfer utilized GUS (β-glucuronidase) as the reporter gene. Saker et al. (2009) reported the first case of successful infection of embryonic cells in date palm and a system was developed for the transfer. Aslam et al. (2015) reported detection of strong GUS activity and integration of *uidA* (*GUS)* and *npt II* genes into transgenic plants in cultivar Khalasah via somatic embryogenesis. Various researchers were involved in the transformation via particle bombardment, but pioneering work was using the Iranian cv. Khorma (Habashi et al. 2008). Likewise, factors affecting transformation via particle bombardment in Egyptian cv. Sewi were optimized (Saker et al. 2007). Mousavi et al. (2014) developed an efficient transformation for gene delivery in date palm. Somatic embryos of Estamaran cultivar were utilized for transient transformation of the uidA gene. Physical and biological parameters were optimized, and tissue bombarded with constructs having the uidA gene. Similarly, for introducing date palm resistance to pests, a construct with the cholesterol oxidase (*ChoA*) gene was introduced into embryonic callus of date palm by particle bombardment (Allam and Saker, 2017).

4 Recent Concepts and Strategies Developed

4.1 Gene Editing

Prevention and treatment of diseases in organisms using genome corrections or modifications refers to gene editing and is gaining ground today as conducted by researchers and scientists. Mainly for therapeutic purposes, gene editing is known

for somatic body cells and germline – reproductive cell-based therapies. The applicability of gene editing is not only limited to animals or humans but also finds an extensive range of applications in plant biotechnology. Gene editing in plants, referring to the alterations in DNA and ultimately affecting the phenotype and associated traits in the plants, involves the use of various enzymes for the insertion and replacement of various DNA sections. Instead of random insertion and deletion of genes like traditional ways in organisms, gene editing targets the site-specific editing at specific locations only and is of immense interest in the treatment and cure of diseases like sickle cell anemia, cystic fibrosis, and others.

The creation of herbicide-resistant, disease-resistant, or nutrient-rich crops is no longer a dream for scientists as they can be constructed even with specific alterations in the plant's own DNA instead of the introduction of new or foreign genes. Both with and without plant tissue culture methods, the creation of identical copies from simple plant tissue is possible with the advancement in technology (Kim 2020). Out of the many techniques available, the two most applied ones in the construction of an edited plant include the addition of *Streptococcus pyogenes* cas9 protein-encoding DNA and a guide RNA using nature's genetic engineer, *Agrobacterium*. CRISPR-Cas9 systems have revolutionized the entire world of plant editing genome and its associated areas. Knock-ins and other methods of delivery for CRISPR and other editing tools require a deeper understanding and knowledge of the processes for finer targeted procedures (Mao et al. 2019). It has been reported that not all CRISPR constructs and variables are successful owing to the differences in specificity required for the expression of Cas9 DNA nuclease, and such a comparison in levels of editing has been exhibited by *Arabidopsis thaliana* (Shockey 2020). Unwanted mutations can also be expected to be observed on introspection at the genomic level as nuclease activity was observed at locations that were targeted using single guide RNAs that were imperfectly matched.

There are processes available that include genetic transformation even without the involvement of DNA, and one of the major concerns includes its cellular genome integration. This often becomes a hurdle in the experimental settings and practical modes, which often drives a reason for developing novel methods of gene editing without the necessity to transform or use DNA delivery (Tsanova et al. 2021). Few of the processes include the formation of ribonucleoprotein complexes, utilization of secretory systems in bacteria for Cas/gRNA delivery, or even application of viruslike particles. For the former option, *Agrobacterium* plays a major role in the CRISPR/Cas9 cassette delivery. Although this method is gaining ground owing to its ease of techniques, low cost, and flexibility, there are a number of obstacles in transformation using genome editing of polyploid crops and germline gene expression (Vats et al. 2019). More precise technologies like promoter bashing, methylome, or gene editing can also be referred to as techniques for advancement in plant sciences and editing.

As far as *Phoenix dactylifera* is concerned, gene editing reports revolve mostly around determination of genetic basicity. Measurable phenotypes as controlled by a set of genes and their interaction with the environment, referred to as quantitative traits, are primarily concentrated on when it comes to gene editing in date palms.

Specific chromosome regions of that control continuous traits are referred to as quantitative trait loci (QTLs). The phenotypic evaluation might not be the ultimate method to identify QTLs, but they can be mapped to identify the location of complex traits as well. In addition, for the identification of QTLs, both genotype and phenotype play a major role by either linkage disequilibrium or overall structure of population or prediction of relatedness in the category of genotypes. CRISPR-associated protein 9-based approaches have been applied in fruit trees; they are cumbersome and practically difficult to apply in date palms due to their higher genome complexity and structure and high outcrossing and heterozygosity rate. The disadvantages of genetic instability owing to high frequency rate of single nucleotide polymorphisms are common in the same. Despite these issues, several approaches using CRISPR/Cas-9 for gene editing in date palms have been enlisted in a report by Sattar et al. (2017). Along with that, various prospects of genome editing tools have also been discussed in date palms.

An array of experiments has been conducted on the genomic content of date palms alone. They have revealed the presence of 38 proteins, 3 ribosomal RNAs, and 30 tRNAs which make up 6.5% of the complete genome and are completely coding, whereas the remaining is noncoding and chloroplast-derived, consisting of tandem and long repeats with a constitution of 0.33% and 2.3%, respectively (Fang et al. 2012). This places the mitochondrial genomic data of date palms at the fourth position based on genomic length. Evolution along parallel lines can be well documented according to a report by Hazzouri et al. (2019) where domesticated species showed diversification in the evolutionary trend. A comprehensive genome assembly of the *P. dactylifera* genome has been made by the group that extends up to 772.3 Mb lengthwise with an 897.2 Kb contig N50, which is further used to execute genome-wide association studies (GWAS) of fruit traits or sex-determining region. It has also been reported that there are 18 polycistronic transcription units and three exclusive genes that are highly expressive in nature, namely *atpF*, *rrn23*, and *trnA-UGC*, as per information from RNA sequencing in date palm chloroplast genome. As observed in most angiosperms, the date palm genome also exhibits 112 unique genes along with 19 duplicated fragments in the inverted repeat regions. There is a typical similarity of date palm genomics with that of tobacco with slight rearrangements in genetic order, as reported by Yang et al. (2010). Major intravarietal polymorphisms as observed include 78 single nucleotide polymorphisms within the chloroplast genome, mostly in genes with vital functions. To get a comprehensive idea about the complications of dioecy and long-time generation, the draft genome for Khalas variety was executed by Al-Dous et al. (2011), constituting a 380 Mb sequence with gene models of more than 25,000, covering 60% of genome and 90% of genes.

4.2 Nanotechnology

The study and application of extremely small objects, being classified on a scale between one and a hundred nanometers on the meter range, refers to the branch of nanotechnology. Currently, it is one of the most developing branches in the

scientific world and rightly gaining prominence among the scientists due to its promising avenues for study of almost any division of research world. The application of nanotechnology has been extended to biological systems especially to create new medicines for combating upcoming challenges in the world of diseases, along with preventing blood clot formation. The most important fluids maintaining the overall physiological conditioning of body, blood, and its components – erythrocytes, leukocytes, and platelets – are accessible to the nanoparticles through the induction of phosphatidylserine on erythrocyte membrane, thereby altering the hemorheological membranes (Zain et al. 2022). An increasing resonance has been found in the application of nanotechnology in the production of secondary metabolites in plants, especially in the usage of silver nanoparticles that possess both hermetic and antimicrobial properties (Rahmawati et al. 2022). This increased applicability is mostly due to microbial decontamination capabilities and elevation in the overall quantity of secondary metabolite content. Detection mechanisms are conducted using nano-biosensors and chips that provide a cheaper, reliable, portable, and faster method for sustainable agricultural procedures (Sellappan et al. 2022). This has already been applied to the detection of microbes, especially viruses in plants as higher pollution was observed on introduction of pesticides in the fields. The never-ending industrialization and urbanization have led to the overall rise in environmental pollution, and application of nanotechnology has helped in the bioremediation processes and decontamination measures taken up by the research centers and industries. In association with nanoparticles, the environmentally friendly and reactivity nature of nanotechnology has helped elevate the quality of bioremediation tasks aided with responsibility in keeping surroundings clean and green. Nyika (2022) has reported various examples where nanotechnology has hastened the process of environmental bioremediation measures that were apparently challenging owing to chemical reactions like photocatalysis, filtration, and other invasive chemical reactions.

For the preparation of nanoparticles, until now several plant extracts have been used as reducing agents, and greener methods of production are more preferred over synthetic ones. Thus, in a report by Batool et al. (2021), nanoparticles were created using *Phoenix dactylifera* as the reducing agent and iron sulfate heptahydrate was the respective substrate to work on. Very interestingly, the nanoparticles have exhibited high antimicrobial activity with a zone of inhibition of about 25 ± 0.360 mm against *Bacillus subtilis*, *Klebsiella pneumoniae*, *Escherichia coli*, and *Micrococcus luteus*. Apart from that, seeds of date palm were used for the same in another experiment by Sirry et al. (2020) for the synthesis of silver nanoparticles. Applications in the biomedical field were checked using the nanoparticles so formed, and extraction was mostly focused on with the usage of several solvents like acidic and alkaline media, ethanol, methanol, and water – both boiling and at normal temperature. In all cases, applicability of water was at a higher efficiency level as compared to other solvents, and anti-inflammatory efficiency was studied by inhibiting albumin denaturation to a higher extent as compared to that using piroxicam alone loaded onto silver nanoparticles.

As explained earlier, almost all the parts of the plant are useful in some way or the other for scientific application. Apart from seeds, leaves of the plant have found applicability in the formation of zinc oxide nanoparticles, where FTIR, TEM, XRD, and UV-visible spectrophotometry are few techniques that have aided in its characterization (Salih et al. 2021). These nanoparticles so formed elevated the physiological characteristics necessitated for growth and development of callus of *Juniperus procera*, as analyzed using GC-MS. Further, there was an upliftment in the quality of biochemical parameters such as chlorophyll a, total flavonoid, and phenol, apart from improvement of total protein and SOD, CAT, and APX activity. In addition, aqueous extract from fruits of date palm have been exclusively used in the preparation of silver nanoparticles and further characterized and evaluated for their in vitro antimicrobial activities (Zafar and Zafar 2019). Potential studies about the cytotoxic activities of the nanoparticles obtained using fruits have been studied in breast cancer cell lines (MCF-7), and cytotoxicity has been found through apoptosis, necrosis, or mechanisms to suppress mitochondrial functioning that can lead to disruptions at various stages in cell cycle. Similar experiments have been performed by Farhadi et al. (2017) using date palm fruits as an extract to produce spherical silver nanoparticles using low-cost and eco-friendly approaches for the preparation of reducing and stabilizing agent from the same. The elemental or crystalline property of nanoparticles was characterized using EDX and XRD techniques, whereas their spherical nature was confirmed using scanning electron microscopy and transmission electron microscopy. Along with that, functional groups as found in biomolecules of date palm fruits are the major reasons for the reduction or stabilization property of nanoparticles, respectively.

There are several active components in *Phoenix dactylifera*, also referred to as date pits, which include L-glutamic acid, gallic acid, or sinapic acid that can function as reducing agents, in accordance with manganese trioxide nanoparticles so synthesized (Sackey et al. 2021). These active compounds have helped in the nanoparticle preparation, and several advanced techniques such as HR-TEM, energy-dispersive detector, high resolution-scanning electron microscopy, and cyclic voltammetry have revealed the efficiency of computationally and experimentally synthesized nanoparticles. Apart from the protocols being cheap, another reason for the preference of green synthesis over synthetic ones includes its environment-friendly nature. Characterization of nanoparticles was conducted using advanced technologies such as UV-visible spectrophotometry showing a peak at 275 nm, indicating the presence of copper oxide. Berra et al. (2018) also confirmed the application of X-ray diffraction and scanning electron microscopy for the crystalline nature and aspherical shape of copper oxide nanoparticles, respectively.

5 Bioinformatics of Date Palm

Bioinformatics is the scientific discipline of collecting and storing biological information in searchable databases and handling the data effectively to draw significant conclusions. Data generated by omics approaches is understood by in silico tools to draw implications about mechanisms of the biological processes in date palm.

To understand the DNA of date palm, whole draft sequence assemblies of nuclear, plastid, and mitochondrial genomes are being generated, and to understand the RNA, protein, phenotype, and chemotype, transcriptomes, proteomes, phenomes, and metabolomes have been generated, respectively.

5.1 Genome Analysis

Date palm is diploid (2n = 36) and several studies have been carried out for the estimation of nuclear genome and organellar genome sizes (Al-Dous et al. 2011; Fang et al. 2012; Al-Mssallem et al. 2013; Sabir et al. 2014a; He et al. 2017; Gros-Balthazard et al. 2018; Thareja et al. 2018; Chaluvadi et al. 2019; Hazzouri et al. 2019; Mohamoud et al. 2019;). Date palm genome assembly draft was analyzed and microsatellite motifs were screened using a script in Perl software, which showed 321,278,327 bases (Hamwieh et al. 2010). Transcriptome sequences were identified and 1000 SSR loci were utilized for developing new markers. 166,760 SSR regions with a density of 1 at every 2.2 Kb in the date palm genome were found (Manju et al. 2016). Xiao et al. (2016) observed at a density of 1 in every 1.36 Kb, 371,629 SSRs. As many as new 264,655 SSR loci were located when 62 cultivar genomes of date palm were re-sequenced and when compared them with known reference genome assemblies available in online database. Pathway annotation work carried out in date palm by using SSR regions and BLAST2GO methods showed 23 enzymes involved in metabolism of starch and sucrose and 16 enzymes involved in metabolism of amino/nucleotide sugar (Mokhtar et al. 2016).

A database has been generated to show the SNP variation, using re-sequenced and known genomes of date palm (He et al. 2017). A density of 1 SNP per 217 bases was seen by date palm genome (Al-Dous et al. 2011). SNP peaks associated with sex-determining loci were found at linkage group (LG) 12 (Mathew et al. 2014; Hazzouri et al. 2019).

5.1.1 Organellar Genome (Chloroplast and Mitochondrial Genome)

Various assembling software like CLC Genomics Workbench (CLC bio, Denmark) were used to assemble the NGS datasets from the pyrosequencing and Illumina systems of date palm chloroplast DNA (Khan et al. 2010). NOVOPlasty program was used for de novo assembly of chloroplast DNA from NGS data along with ORF Finder (http://www.ncbi. nlm.nih.gov/projects/gorf/), and Dual Organellar Genome Annotator (the DOGMA server) was employed for the annotation of chloroplast genomes (Wyman 2004; Dierckxsens et al. 2017; Rasheed et al. 2020). Also, tRNAscan-SE was carried out for annotation of some tRNAs (Lowe and Eddy 1997), and description of repeat sequences was carried out by using the REPuter program after searching for sequence similarity with annotated plastomes (Kurtz et al. 2001). Bioinformatics tools including the GenomeVx online and Gene Order tool were used for the circular genome map of date palm, and chloroplast genome and gene order were investigated (Celamkoti 2004; Conant 2008). mVISTA comparative genomics server was used for the construction of multiple sequence alignments of chloroplast genomes (Frazer et al. 2004). MEGA4 program was utilized for

parsimony-based phylogenetic tree construction (Tamura et al. 2007). Various bioinformatics software had been used to examine and infer the date palm cp DNA sequence, and it showed a close association with the broad leaf cattail (*Typha latifolia* L.). The mitochondrial genomes of date palm are the circular DNAs of sizes ranging from 585,493 bp to 715,120 bp. In *Phoenix dactylifera* (NC016740), *P. dactylifera* (MH176159), and *P. dactylifera* cv. Naghal, the protein-coding sequence is composed of 5.35, 4.6, and 4.6% gene content, respectively. However, the noncoding region possesses only 1.8% in the unverified mitochondrial genome of *P. dactylifera* (MG257490). The gene content is similar among these mt genomes. In *P. dactylifera* (NC016740), in total 43 protein-coding genes were found.

5.1.2 Whole Genome

Numerous web-based date palm databases have been constructed to map the date palm genome. Most databases available today are intended to help researchers and breeders to recognize predominant date palm cultivars by selecting highly specific polymorphic markers. These databases help researchers to search genetic variations among different date palm cultivars, thus helping in further breeding and genetic studies. The following are the databases used to map the date palm genome.

Date Palm Genome Database (DRDB)

DRDB was developed as a platform to help researcher and breeder in observing and recognizing different date palm varieties or cultivars using polymorphic markers using search functionalities that can be accessed freely – http://drdb.big.ac.cn/home (He et al. 2017). The database has sequenced genomic data of 62 different cultivars from different parts of the world, with 246,445 SSRs and 6,375,806 SNPs annotated in the genome assembly. Phylogenetic tree developed from these cultivars based on the geographic location shows three subclades. The database also helps in identifying and retrieving the SSR and SNP data sheets of specified cultivars, which can be customized based on requisite parameters. In addition, markers can be selected at cultivar, regional, or country level. The database also provides information associated with SNP annotation, etc. Information for external sources like UniProt is also available. The data can be accessed freely and downloaded.

Plant Genome and System Biology (PGSB)

PGSB is an online database comprising the databases of various plants. The date palm genome database can be accessed directly at http://pgsb.helmholtz-muenchen.de/plant/pdact/ index.jsp or https://qatar-weill.cornell.edu/ research/research-highlights/date-palm-research program. Date palm genome sequence and genome annotation is available at date palm draft sequence version 3.0. Shotgun next-generation DNA sequencing was utilized for creation of date palm genome draft assembly with an estimated genome size of ca 650 Mb. Over 3.5 million high-quality SNPs distributed among nine date palm genomes and the Khalas cv. as a reference genome is available. 90% of genes and 60% of the genome sequences including repetitive sequences and the chloroplast genome account for 381 MB of assemble sequences. Contig tables and genetics elements are placed as separate sections which can be accessed by name/id or free text methods. Information is available freely and can be

downloaded from http://qatar-weill.cornell.edu/research/research-highlights/date-palm-researchprogram/datepalm-draft-sequence.

Date Palm Molecular Markers Database (DPMMD)

DPMMD contains information useful for basic and applied research that can be accessed freely at http://dpmmd. easyomics.org/index.php. DPMMD provides information of more than 3,611,400 DNA markers in addition to genetic linkage maps, KEGG pathways, DNA barcodes, and date palm markers related to articles indexed in PubMed journals. In DPMMD, a list of SSR and SNP primers is integrated from some previously published data (Al-Dous et al. 2011; Mokhtar et al. 2016) and compiled in a relational database using MySQL environment. Information of SSR markers is categorized into Genic SSR, Genic SSR-SNP, SNP markers, intergenic SSR-SNP and R-gene SSR markers, and DNA barcode. In addition, it has Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway where database 896 SSR markers were mapped to 111 pathways. DPMMD also contains the first ever constructed date palm genetic linkage maps of 1293 cM distance.

NGS-Based Sequencing

Gene functions and their involvement in pathways of date palm were determined by NGS-based sequencing techniques coupled with omics technologies (Safronov et al. 2017). The prospects of genetic mapping as a complementary method for the identification of QTLs related to abiotic stress have also been identified (Hazzouri et al. 2020).

5.2 Gene Annotation and Promoter Motifs

By using bioinformatics approaches, abiotic stress-responsive genes were analyzed. Bioinformatics methods were used to find the desaturases of date palm and their role in stress response (Sham and Aly 2012) and motif abundances in abiotic (heat, drought) stress-induced genes (Safronov et al. 2017). In aquaporin gene PdPIP1, two regions of date palm indicated abundant abiotic (drought and salinity) stress-inducible motifs in its 2 Kb promoter region and occurrence of four groups of aquaporins in date palm genome in 40 members (Patankar et al. 2019). Phytoremediation by date palm of polluted sites to remove toxic metals especially Cd, Cu, Pd, and Cr and harmful dioxins has shown promising results (Hanano et al. 2016; Al-Najar et al. 2019; Sivarajasekar et al. 2019). Phytochelatin synthase (PdPCS1) was investigated in silico distributed in eight exons intervened by seven introns and was about 5.7 Kb long (Zayneb et al. 2017).

5.3 Gene Mapping for Trait-Linked Attributes

5.3.1 Sex-Linked Attributes

Several studies were carried out to determine the sex of date palm seedlings by using markers (Heikrujam et al. 2015; Awan et al. 2017). XY sex determination pattern is seen in date palm. The sex-linked SRY region which is specific to male plants

possesses recombination arrest mechanism. Analysis using bioinformatic tools was carried out to understand the gene, SSR, and SNP variation using transcriptomes and whole genomes of date palm (Hamweih et al. 2010; Al-Dous et al. 2011; Torres et al. 2018; Arunachalam 2021) to differentiate between sexes. The sex-determining region of the date palm having a size of approximately 6–13 Mb is present in the telomeric region of the long arm of linkage group (LG) (Mathew et al. 2014; Cherif et al. 2016; Hazzouri et al. 2019). SSR markers, SCAR marker, etc. have been developed for the identification of sexes (Billotte et al. 2004; Arunachalam 2021). The physical mapping of the genes and genomic regions on date palm LG 12 chromosome was developed by Mathew et al. (2014) to understand genomic arrangement and the evolution of sex by *in silico* methods (Premkrishnan and Arunachalam 2012; Thiel et al. 2003).

5.3.2 Attributes

Fruit color differs in different cultivars of date palm. For example, dominant negative mutation suppresses purple color anthocyanin pigment to cause yellow peel color of the date fruit (Hazzouri et al. 2019). The pericarp color of the date fruit is determined by R2-R3-MYB transcription factor (VIR locus) virescence gene, etc. (Hazzouri et al. 2015). Similarly, change in color from purple to yellow is caused by retrotransposon insertion of 397 bp size in the exon region leading to truncation of R2-R3-MYB gene (Hazzouri et al. 2019). Deletion of copies of invertase gene can lead to changes in the content of sucrose, fructose, and glucose among cultivars (Ghnimi et al. 2017; Hazzouri et al. 2019; Malek et al. 2020).

5.4 MicroRNA Prediction

Gene expression at the posttranscriptional stage is regulated by microRNA (miRNA) molecules. An eightfold increase in galactose content in heat- and drought-stressed date palm in metabolome of date palm has been observed (Safronov et al. 2017).

5.5 Image Analysis and Molecular Structure Analysis

Fluorescence microscopy is used for phenotyping of date palm varieties like leaflet anatomy (Arinkin et al. 2014), size, and shape of the seed (Terral et al. 2012). The phytochelator synthase gene structure of date palm was predicted using the available structure of a similar gene from the cyanobacteria genus *Nostoc* as a template (Zayneb et al. 2017).

6 Conclusion and Prospects

Date fruit is highly nutritious compared to any other fruits. Several varieties of date fruits can be consumed which are associated with any metabolic syndrome and related diseases. Dates being a good source of antioxidants have a vital role in

regulating human health. Hence, date fruit can be considered as a remarkable medicinal fruit due to its nutritive, bioactive, and therapeutic potentials. But there is a strong need to explore the nutra-pharmaceutical and health benefits of dates on the basis of their functional components and elucidate the mechanisms of action of such bioactives.

The available genomic resources of date palm are useful for finding out bioinformatics approaches to understand the genomic variations among cultivars in different geographical regions. Availability of the whole genome sequence, organellar sequence, and genetic map of date palm have helped breeders to modify and improve various agronomic characteristics by manipulating the genome, to improve nutritional components of the fruits. The date palm genome provides scope to understand several other biological processes with unique features. Genes and markers linked to iron content in fruit in different cultivars of date palm need to be explored. Identified SNP markers till now in nuclear and organelle genomes of date palm provide potential for developing SNP chips for molecular marker-assisted selection of date palm. Molecular modeling and docking studies will improve our understanding of the gene structure, binding sites, and crucial residues. The data on the status of research by material and software used for in silico studies on date palm omics provided here can help to identify drawbacks, to further plan investigation on the nutritional value and iron content of fruits in different varieties of date palm, to develop new cultivars, and to promote the consumption and utilization of date palm worldwide in place of several staple food used at present.

Acknowledgments The authors express their appreciation for the support provided by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. GRANT3173].

References

Abass KI, Hameed EH, Al-Ramadan AAA (2012) Effect of chelated iron and benzyl-adenine on physiochemical characteristics and yield of date palm (*Phoenix dactylifera* L.) CV "Hellawi". J King Abdulaziz Univ Meteorol Environ Arid Land Agric Sci 23(1):133

Abbas KI, Saad AA (2009) Effect of some amino acids, vitamins and plant growth regulators on maintenance and length of plantlets derived from in vitro culture of date palm (*Phoenix dactylifera* L.). CV. Barhee. J Kerbala Univ 7(1):54–60

Abbas KI, Taain DA, Wahid AM (2007) Study the effect of nitrogen and iron application on productivity of date palm *Phoenix dactylifera* L. cv. Hellawi. Jo Basrah Res (Sciences) 33(3B): 15–19

Abd Elwahab SM, Abd Allatif AM, Farid MA, Soliman SM (2019) Effect of safe post-harvest alternatives on quality and storage life of "barhi" date palm. Plant Arch 19(2):3937–3945

Abdalla RSM, Eltayeb MAH, Hassan AME, El Hussein AM (2020) Determinations of amino acids and fatty acids levels in date palm kernel flour in Sudan. J Agric Sci Technol. https://doi.org/10.17265/2161-6256/2020.05.000

Abdul-Hamid NA, Abas F, Ismail IS, Tham CL, Maulidiani M, Mediani A, Swarup S, Umashankar S (2019) 1H-NMR-based metabolomics to investigate the effects of *Phoenix dactylifera* seed extracts in LPS-IFN-γ-induced RAW 264.7 cells. Food Res Int 125:108565

Abdulla M, Gamal O (2010) Investigation on molecular phylogeny of some date palm (*Phoenix dactylifera* L.) cultivars by protein, RAPD and ISSR markers in Saudi Arabia. Aust J Crop Sci 4: 23–28

Abul-Soad AA, Mahdi SM, Markhand GS (2015) Date palm status and perspective in Pakistan. In: Al-Khayri JM, Jain SM, Johnson DV (eds) Date palm genetic resources and utilization, vol 2. Asia and Europe. Springer, Dordrecht, pp 153–205

Abul-Soad AA, Jain SM, Jatoi MA (2017) Biodiversity and conservation of date palm. Biodiversity and conservation of woody plants 313–353

Acharya A, Pal PK (2020) Agriculture nanotechnology: translating research outcome to field applications by influencing environmental sustainability. NanoImpact 19:100232

Adawy SS, Ebtissan HA, Hussein D, El-Khishin D, Saker Mohammed AA, El-Itriby HA (2004) Genotyping Egyptian date palm cultivars using RAPD, ISSR and AFLP markers and estimation of genetic stability among tissue culture derived plants. In: Bibliotheca Alexandria Conference Center, Alexandria, Egypt

Adenekan O, Omoyajowo K, Babalola O, Amiolemen JOS, Olaniyan K, Akande J, Idowu I (2018) Comparative nutritional analysis of *Phoenix dactylifera* and *Phoenix reclinata* seeds. J Res Rev Sci 4:29–35

Agbon AN, Kwaneshie HO, Hamman WO (2013) Antidiarrheal activity of aqueous fruit extract of *Phoenix dactylifera* (DATE PALM) in Wistar rats. Br J Pharmacol Toxicol 4(3):121–127

Agwa MM (2021) Novel transcriptome study and detection of metabolic variations in UV-B-treated date palm (*Phoenix dactylifera* cv. Khalas). Int J Mol Sci 22(5):2564.

Ahmad M, Akanji MA, Usman AR, Al-Farraj AS, Tsang YF, Al-Wabel MI (2020) Turning date palm waste into carbon nanodots and nano zerovalent iron composites for excellent removal of methylthioninium chloride from water. Scientific reports 10(1):16125

Ahmad Mohd Zain MR, Abdul Kari Z, Dawood MA, Nik Ahmad Ariff NS, Salmuna ZN, Ismail N, Ibrahim AH, Thevan Krishnan K, Che Mat NF, Edinur HA, Abdul Razab MK (2022) Bioactivity and pharmacological potential of date palm (*Phoenix dactylifera* L.) against pandemic COVID-19: a comprehensive review. Appl Biochem Biotechnol 17:1–38

Ahmed TA, Al-Qaradawi AYK (2009) Molecular phylogeny of Qatari date palm genotypes using simple sequence repeats markers. Biotechnology 8:126–131

Ahmed M, Soliman S, Elsayed E (2006) Molecular identification of some Egyptian date palm males by females varieties (*Phoenix dactylifera* L.) using DNA markers. J Appl Sci Res 2:270–275

Ahmed MVOM, Bouna ZEO, Lemine FMM et al (2011) Use of multivariate analysis to assess phenotypic diversity of date palm (*Phoenix dactylifera* L.) cultivars. Sci Hortic 127(3):367–371

Akkak A, Scariot V, Marinoni DT, Boccacci P, Beltramo C, Botta R (2003) Development and evaluation of microsatellite markers in *Phoenix dactylifera* L. and their transferability to other *Phoenix* species. Biol Plant 53:164–166

Al Alawi R, Alhamdani MSS, Hoheisel JD, Baqi Y (2020) Antifibrotic and tumor microenvironment modulating effect of date palm fruit (*Phoenix dactylifera* L.) extracts in pancreatic cancer. Biomed Pharmacother 121:109522

Al Juhaimi F, Özcan MM, Uslu N, Ghafoor K, Babiker EE, Mohamed Ahmed IA (2020) Bioactive properties, fatty acid compositions, and phenolic compounds of some date palm (*Phoenix dactylifera* L.) cultivars. J Food Process Preserv 44(5):e14432

Al Ohali Y (2011) Computer vision-based date fruit grading system: design and implementation. J King Saud Univ-Comput Inf Sci 23(1):29–36

Alam F, Rahman MS, Alam MS, Hossain MK, Hossain MA, Rashid MA (2009) Phytochemical and biological investigations of *Phoenix paludosa* Roxb. Dhaka Univ J Pharm Sci 8(1):7–10

Al-Ameri AA, Al-Qurainy F, Gaafar ARZ et al (2016a) Molecular identification of sex in *Phoenix dactylifera* using inter simple sequence repeat markers. Biomed Res Int 2016:4530846

Al-Ameri AA, Al-Qurainy F, Gaafar ARZ et al (2016b) Male specific gene expression in dioecious *Phoenix dactylifera* (date palm) tree at flowering stage. Pak J Bot 48(1):131–135

Al-Doss AA, Aly MA, Bacha MA (2001) Morphological and agronomical variations among some date palm cultivars grown in Saudi Arabia using principal component and cluster analysis. King Saud Univ Agric J 13:3–18. (In Arabic)

Al-Dous EK, George B, Al-Mahmoud ME, Al-Jaber MY, Wang H, Salameh YM, Al-Azwani EK, Chaluvadi S, Pontaroli AC, Debarry J, Arondel V, Ohlrogge J, Saie IJ, Suliman-Elmeer KM, Bennetzen JL, Kruegger RR, Malek JA (2011) De-novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). Nat Biotechnol 29:521–528

Alem C, Ennassir J, Benlyas M, Mbark AN, Zegzouti YF (2017) Phytochemical compositions and antioxidant capacity of three date (*Phoenix dactylifera* L.) seeds varieties grown in the South East Morocco. J Saudi Soc Agric Sci 16(4):350–357

Al-Faifi SA, Migdadi HM, Algamdi SS, Khan MA, Al-Obeed RS, Ammar MH, Jakse J (2017) Development of genomic simple sequence repeats (SSR) by enrichment libraries in date palm. Methods Mol Biol 1638:315–337

Alfei S, Marengo B, Zuccari G (2020) Nanotechnology application in food packaging: a plethora of opportunities versus pending risks assessment and public concerns. Food Res Int 137:109664

Alghamdi SB, Togoo RA, Bahamdan GK, Delaim A, Asiri EA, Mallawi GZ, Alkahtani ZM (2019) Changes in salivary pH following consumption of different varieties of date fruits. J Taibah Univ Med Sci 14(3):246–251

Al-Hajaj HS, Ayad JY, Othman YA, Abu-Rayyan A (2020) Foliar potassium application improves fruits yield and quality of 'Medjool'date palm. Fresenius Environ Bull 29(3):1436–1442

Aljaloud S, Colleran HL, Ibrahim SA (2020) Nutritional value of date fruits and potential use in nutritional bars for athletes. Food and Nutrition Sciences 11(6):463

Al-Khayri JM, Naik PM, Jain SM, Johnson DV (2018) Advances in date palm (Phoenix dactylifera L.) breeding. In J. M. Al-Khayri, S. M. Jain, & D. V. Johnson (Eds.), Advances in Plant Breeding Strategies: Fruits: vol. 3 (pp. 727–771). Springer International Publishing. https://doi.org/10.1007/978-3-319-91944-7_18

Al-Khalifah NS, Askari E (2003) Molecular phylogeny of date palm (*Phoenix dactylifera* L.) cultivars from Saudi Arabia by DNA fingerprinting. TheorAppl Genet 107:1266–1270

Al-Khalifah NS, Askari E (2006) Early detection of genetic variation in date palms propagated from tissue culture and offshoots by DNA fingerprinting. In III International Date Palm Conference 736:105–112

Al-Khalifah NS, Askari E, Khan AES (2012) Molecular and morphological identification of some elite varieties of date palms grown in Saudi Arabia. Emir J Food Agric 24(5):456–461

Al-Khateeb TA, Jubrael JM (2006) The use of RAPD markers for sex and male cultivars identification in (*Phoenix dactylifera* L.) in Iraq. Acta Hortic 736:162–170

Al-Khayri JM, Al-Bahrany AM (2004) Growth, water content, and proline accumulation in drought-stressed callus of date palm. Biol Plant 48(1):105–108

Al-Khayri JM, Jain SM, Johnson DV (2021) The date palm genome, vol 2: Omics and molecular breeding (compendium of plant genomes). Springer International Publishing, Cham

Allam MA, Saker MM (2017) Microprojectile bombardment transformation of date palm using the insecticidal cholesterol oxidase (ChoA) gene. In: Date palm biotechnology protocols, vol I. Humana Press, New York, pp 281–293

Al-Mahmoud ME, Al-Dous EK, Al-Azwani EK et al (2012) DNA-based assays to distinguish date palm (Arecaceae) gender. Amer J Bot 99(1):7–10

Al-Moshileh AM, Motawei MI, Al-Wasel A, Abdel-Latif T (2004) Identification of some date palm (*Phoenix dactylifera* L.) cultivars in Saudi Arabia using RAPD fingerprints. Agri Marine Sci 9:1–3

Al-Mssallem IS, Hu S, Zhang X et al (2013) Genome sequence of the date palm *Phoenix dactylifera* L. Nat Commun 4:2274

Al-Najar B, Bououdina M, Vijaya JJ et al (2019) Removal of toxins from the environment using date palm seeds. In: Naushad M, Lichtfouse E (eds) Sustainable agriculture reviews, vol 34. Springer, Cham, pp 207–245

Al-Qarawi AA, Abdel-Rahman H, Ali BH, Mousa HM, El-Mougy SA (2005) The ameliorative effect of dates (*Phoenix dactylifera* L.) on ethanol-induced gastric ulcer in rats. J Ethnopharmacol 98(3):313–317

Al-Qarawi AA, Abdel-Rahman H, Mousa HM, Ali BH, El-Mougy SA (2008) Nephroprotective action of *Phoenix dactylifera* in gentamicin-induced nephrotoxicity. Pharm Biol 46(4):227–230

Al-Qurainy F, Al-Ameri AA, Khan S et al (2018) SCAR marker for gender identification in date palm (*Phoenix dactylifera* L.) at the seedling stage. Int J Genom 2018:3035406

Al-Samarai AH, Al-Salihi FG, Al-Samarai RR (2018) Phytochemical constituents and nutrient evaluation of date palm (*Phoenix dactylifera*, L.) pollen grains. Tikrit J Pure Sci 21(1):56–62

Al-Sayyed HF, Abu-Qatouseh LF, Malkawy M, Al-Wawi S, Al Kafaween M (2021) Extracts of Jordanian date palm fruit (*Phoenix dactylifera* L.) inhibit human mammary adenocarcinoma (MCF-7) cells in vitro by inducing cell viability. Curr Res Nutr Food Sci J 9(2):423–430

Al-Shahib W, Marshall RJ (2003a) Fatty acid content of the seeds from 14 varieties of date palm *Phoenix dactylifera* L. Int J Food Sci Technol 38(6):709–712

Al-Shahib W, Marshall RJ (2003b) The fruit of the date palm: its possible use as the best food for the future? Int J Food Sci Nutr 54(4):247–259

Altememe I, Mahdi AS (2015) The effects of fertilizer (Fetrilon Combi2) spraying and dates of application on chemical characteristics and yield of the date palm *Phoenix dactylifera* L. cv. Hillawi. Basrah J Date Palm Res 11(1):1–24

Altememe ZAM, Ali AA, Shareef MH, Kadhim AQ (2017) The physiological study of the effect of foliar spray of humic acid and compound fertilizers N.P.K on some chemical characteristics of date palm cultivar maktoom in the city of Karbala. Karbala J Agric Sci 4(1):1–13

Al-Wasfy MM (2013) Response of Sakkoti date palms to foliar application of royal jelly, silicon and vitamins B. J Am Sci 9(5):315–321

Al-Whaibi MH, Basalah MO, Al-Achal IE (1985) Chemical Composition of Datepalm (*Phoenix dactylifera* L.), J Coll Sci King Saud Univ 16(10):23

Al-Yahyai R, Al-Khanjari S (2008) Biodiversity of date palm in the Sultanate of Oman. Afr J Agric Res 3:389–395

Al-Yahyai R, Khan MM (2015) Date palm status and perspective in Oman. In: Al-Khayri JM, Jain SM, Johnson DV (eds) Date palm genetic resources and utilization, Asia and Europe, vol 2. Springer, Dordrecht, pp 207–240

Alyileili SR, Hussein AS, Ibrahim W, El-Tarabily KA (2020) Phytochemical composition and antioxidant activity of *Trichoderma reesei* degraded date (*Phoenix dactylifera* L.) pits. Curr Bioact Compd 16(4):528–536

Arinkin V, Digel I, Porst D et al (2014) Phenotyping date palm varieties via leaflet cross-sectional imaging and artificial neural network application. BMC Bioinform 15(1):55

Arunachalam V (2021) Date palm bioinformatics. In: The date palm genome, vol 1. Springer, Cham, pp 223–235

Asif MI, Al-Taher OA (1983) Ripening of Khasab dates by sodium chloride and acetic acid. Date Palm J 2:121–128

Aslam J, Khan SA, Azad MAK (2015) Agrobacterium-mediated genetic transformation of date palm (*Phoenix dactylifera* L.) cultivar "Khalasah" via somatic embryogenesis. Plant Sci Today 2 (3):93–101

Assirey EAR (2015) Nutritional composition of fruit of 10 date palm (*Phoenix dactylifera* L.) cultivars grown in Saudi Arabia. J Taibah Univ Sci 9(1):75–79

Ataei D, hamidi-Esfahani Z, Ahmadi-Gavlighi H (2020) Enzymatic production of xylo-oligosaccharide from date (*Phoenix dactylifera* L.) seed. Food Sci Nutr 8(12):6699–6707

Atia MAM, Sakr MM, Mokhtar MM, Adawy SS (2017) Development of sex-specific PCR-based markers in date palm. Methods Mol Biol 1638:227–244

Attaha AH, Al-Mubark NR (2014) Effect of spraying seaweed extract kelpak and NPK fertilizer on nitrogen, phosphorus, potassium and total protein concentrations of leaves and fruits of *Phoenix dactylifera* L., cv. Barhi. Jo Basrah Res (Sciences) 40(1B):65–84

Attia AI, Reda FM, Patra AK, Elnesr SS, Attia YA, Alagawany M (2021) Date (*Phoenix dactylifera* L.) by-products: chemical composition, nutritive value and applications in poultry nutrition, an updating review. Animals 11(4):1133

Auda H, Mirjan J, Al-Wandawi H, Al-Charchafchi F (1973) Chemical changes induced by radiation in Iraqi dates. In: Radiation preservation of food, Vienna (Austria), Symposium on radiation preservation of food. International Atomic Energy Agency, Vienna (Austria); Food and Agriculture Organization of the United Nations, Rome (Italy); Proceedings series; p. 21–27

Auda H, Al-Wandawi H, Al-Adhami L (1976) Protein and amino acid composition of three varieties of Iraqi dates at different stages of development. J Agric Food Chem 24(2):365–367

Awad MA (2007) Increasing the rate of ripening of date palm fruit (*Phoenix dactylifera* L.) cv. Helali by preharvest and postharvest treatments. Postharvest Biol Technol 43(1):121–127

Awan FS, Jaskani MJ, Sadia B (2017) Gender identification in date palm using molecular markers. In: AlKhayri JM, Jain SM, Johnson DV (eds) Date palm biotechnology protocols, vol II. Humana Press, New York, pp 209–225

Awofadeju OFJ, Awe AB, Adewumi OJ, Adeyemo EA (2021) Influence of substituting sucrose with date palm fruit flour (DPFF) on the nutritional and organoleptic properties of bread. Croatian J Food Sci Technol 13(1):1–6

Ba-Angood S (2015) Date palm status and perspective in Yemen. In: Al-Khayri JM, Jain SM, Johnson DV (eds) Date palm genetic resources and utilization. vol 2: Asia and Europe. Springer, Dordrecht, pp 241–263

Bacigalupa ZA, Bhadiadra CH, Reginato MJ (2018) O-GlcNAcylation: key regulator of glycolytic pathways. J Bioenerg Biomembr 50(3):189–198

Badarusham K, Sabri NE, Salvamani S, Hassan MS, Hassan Z, Hashim R (2019) Assessment of minerals in *Phoenix dactylifera* L. as determined by inductively coupled plasma optical emission spectrometry using ANOVA and PCA. Int J Eng Technol

Barakat AZ, Hamed AR, Bassuiny RI, Abdel-Aty AM, Mohamed SA (2020) Date palm and saw palmetto seeds functional properties: antioxidant, anti-inflammatory and antimicrobial activities. J Food Meas Charact 14(2):1064–1072

Barreveld WH (1993) Date palm products. Agricultural services bulletin 101. Food and Agriculture Organisation of the United Nations, Rome

Barrow SA (1998) Monograph of *Phoenix* L. (Palmae: Coryphoideae). Kew Bull 53:513–575

Bashah MA (1996) Date variety in the Kingdom of Saudi Arabia. Guide booklet for palms and dates. King Abdul Aziz Univ Press, Jeddah, pp 1225–1319

Batool F, Iqbal MS, Khan SUD, Khan J, Ahmed B, Qadir MI (2021) Biologically synthesized iron nanoparticles (FeNPs) from *Phoenix dactylifera* have anti-bacterial activities. Sci Rep 11(1):1–9

Battaglia M, Ghsera B, Mancini M, Bergesio C, Camussi A, Racchi ML (2015) Date palm status and perspective in Libya. In: Al-Khayri JM, Jain SM, Johnson DV (eds) Date palm genetic resources and utilization, vol 1. Africa and the Americas. Springer, Dordrecht, pp 223–255

Bekheet S, Taha H, Hanafy M et al (2008) Morphogenesis of sexual embryos of date palm cultured in vitro and early identification of sex type. J Appl Sci Res 4:345–352

Ben-Abdallah A, Stiti K, Lepoivre P, du Jardim P (2000) Date palm (*Phoenix dactylifera* L.) cultivar identification using random amplified polymorphic DNA (RAPD). Cah Agric 9:103–107

Benkhalifa A (1999) Gestion de la diversitégénétique du palmier dattierenAlgérie. Paper presented at the workshop constitution etorganisationd'équipes de recherchescientifique dans les domaines de foresterie et des arbresfruitiers, Marrakech, 13–15 Apr

Bennett RD, Ko ST, Heftmann E (1966) Isolation of estrone and cholesterol from the date palm, *Phoenix dactylifera* L. Phytochemistry 5(2):231–235

Bensaci C, Ghiaba Z, Dakmouche M, Belfar A, Belguidoum M, Bentebba FZ, Saidi M, Hadjadj M (2021) In vitro evaluation of antioxidant potential of date palm collected in Algeria using electrochemical and spectrophotometrical techniques. Korean Chem Eng Res 59(2):153–158

Berra D, Laouini SE, Benhaoua B, Ouahrani MR, Berrani D, Rahal A (2018) Green synthesis of copper oxide nanoparticles by *Phoenix dactylifera* L leaves extract. Digest J Nanomater Biostructures 13(4):1231–1238

Billotte N, Marseillac N, Brottier P, Noyer JL, Jacquemoud-Collet JP, Moreau C, Couvreur T, Chevallier MH, Pintaud JC, Risterucci AM (2004) Nuclear microsatellite markers for the date palm (Phoenix dactylifera L.): characterization and utility across the genus Phoenix and in other palm genera. Molecular Ecology Notes 4(2):256–258

Bijami A, Rezanejad F, Oloumi H, Mozafari H (2020) Minerals, antioxidant compounds and phenolic profile regarding date palm (*Phoenix dactylifera* L.) seed development. Sci Hortic 262:109017

Bodian A, NdoyeNdir K, Diedhiou PM, Borgel A, Sane D, Sagna M (2012a) Analysis of genetic diversity of date palm (*Phoenix dactylifera* L.) cultivars from Mauritania using microsatellite markers. Int J Sci Adv Technol 2:36–43

Bodian A, Elhoumaizi MA, Ndoye Ndir K, Hasnaoui A, Nachtigall M, Wehling P (2012b) Genetic diversity analysis of date palm cultivars from Figuig oasis (Morocco) using SSR markers. Int J Sci Adv Technol 2:96–104

Boudeffeur S, Ameur AA, Aci MM, Khelafi H, Laaboudi A, Khersi M, . . .& Khelifi L (2021) Phenotypic and genetic characterization of date palm cultivars resistant to bayoud disease. Plant Sci Today 8(4):804–814

Bouguedoura N, Bennaceur M, Babahani S, Benziouche SE (2015) Date Palm Status and Perspective in Algeria. In: Al-Khayri, J., Jain, S., Johnson, D. (eds) Date Palm Genetic Resources and Utilization. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-9694-1_4

Cadi HE, Bouzidi HE, Selama G, Ramdan B, Majdoub YOE, Alibrando F, Arena K, Lovillo MP, Brigui J, Mondello L, Cacciola F, Salerno TM (2021) Elucidation of antioxidant compounds in Moroccan *Chamaerops humilis* L. fruits by GC–MS and HPLC–MS techniques. Molecules 26 (9):2710

Casagrande E, Génard M, Lurol S, Charles F, Plénet D, Lescourret F (2021) A process-based model of nectarine quality development during pre-and post-harvest. Postharvest Biol Technol 175: 111458

Celamkoti S (2004) GeneOrder 3.0: software for comparing the order of genes in pairs of small bacterial genomes. BMC Bioinform 5:52

Chaluvadi SR, Young P, Thompson K et al (2019) Phoenix phylogeny, and analysis of genetic variation in a diverse collection of date palm (*Phoenix dactylifera*) and related species. Plant Diver 41(5):330–339

Chaudhary S, Kumar S, Kumar V, Sharma R (2020) Chitosan nanoemulsions as advanced edible coatings for fruits and vegetables: composition, fabrication and developments in last decade. Int J Biol Macromol 152:154–170

Cherif E, Zehdi-Azouzi S, Crabos A et al (2016) Evolution of sex chromosomes prior to speciation in the dioecious Phoenix species. J Evol Biol 29(8):1513–1522

Conant GC (2008) Genome Vx: simple web-based creation of editable circular chromosome maps. Bioinform 24:861–862

Cornicquel B, Mercier L (1994) Date palm (*Phoenix dactylifera* L.) cultivar identification by RFLP & RAPD. Plant Sci 101:163–172

Cornicquel B, Mercier L (1997) Identification of date palm (*Phoenix dactylifera* L.) cultivars by RFLP: partial characterization of a cDNA probe that contains a sequence encoding a zinc finger motif. Int J Plant Sci 158:152–156

Das S, Acharya J, De B (2017) Metabolite profiling, antioxidant activity, and glycosidase inhibition property of the mesocarp tissue extracts of sugar date palm [*Phoenix sylvestris* (L.)Roxb.] fruits. Int J Food Prop 20(12):2982–2993

De Jager A, Roelofs FPMM (1996) Prediction of optimum harvest date of Jonagold. Determination and prediction of optimum harvest date of apples and pears. COST 94:21–31

Dhawan C, Kharb P, Sharma R et al (2013) Development of male-specific SCAR marker in date palm (*Phoenix dactylifera* L.). Tree Genet Genomes 9:1143–1150

Dhawi F, Al-Khayri JM (2008) Proline accumulation in response to magnetic fields in date palm (*Phoenix dactylifera* L.). Open Agric J 2(1):80–83

Di Cagno R, Filannino P, Cavoski I, Lanera A, Mamdouh BM, Gobbetti M (2017) Bioprocessing technology to exploit organic palm date (*Phoenix dactylifera* L. cultivar Siwi) fruit as a functional dietary supplement. J Funct Foods 31:9–19

Diab KAS, Aboul-Ela E (2012) In vivo comparative studies on antigenotoxicity of date palm (*Phoenix dactylifera*L.) pits extract against DNA damage induced by N-Nitroso-N-methylurea in mice. Toxicol Int 19(3):279

Diaz S, Pire C, Ferrer J, Bonete MJ (2003) Identification of *Phoenix dactylifera* L. varieties based on amplified fragment length polymorphism (AFLP) markers. Cell Mol Biol Lett 8:891–899

Díaz-Gómez J, Twyman RM, Zhu C, Farré G, Serrano JC, Portero-Otin M, Muñoz P, Sandmann G, Capell T, Christou P (2017) Biofortification of crops with nutrients: factors affecting utilization and storage. Curr Opin Biotechnol 44:115–123

Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res 45(4):e18

Djihad, T., Alghorbany, A., Alam, M., & Bin Muhamad, A. I. (2021). Government Policies, Financial Scopes and Technological Usages for Agricultural Development and Post-Harvest Loss Reduction in Algeria. Int J Postharvest Technol Innov 7(4):335–352

Dole John M, Faust James E (2021) Post-Harvest Management. In: Cut flowers and foliages, p 342, CAB International.

Eeuwens CJ (1978) Effects of organic nutrients and hormones on growth and development of tissue explants from coconut (*Cocos nucifera*) and date (*Phoenix dactylifera*) palms cultured in vitro. Physiol Plant 42(2):173–178

Eid N, Enani S, Walton G, Corona G, Costabile A, Gibson G, Rowland I, Spencer JP (2014) The impact of date palm fruits and their component polyphenols, on gut microbial ecology, bacterial metabolites and colon cancer cell proliferation. J Nutr Sci 3:e46

Eissa EA, Abdelrazak AB, Elsharabishi SF, Rizk RM (2009) Morphological and molecular genetic characterization of soft date palm (*Phoenix dactylifera* L.) cultivars in Egypt. Egypt J Genet Cytol 38:269–284

El-Assar AM, Krueger RR, Devanand PS, Chao CT (2005) Genetic analysis of Egyptian date (*Phoenix dactylifera* L.) accessions using AFLP markers. Genet Resour Crop Evol 52:601–607

Elberry AA, Mufti ST, Al-Maghrabi JA, Abdel-Sattar EA, Ashour OM, Ghareib SA, Mosli HA (2011) Anti-inflammatory and antiproliferative activities of date palm pollen (*Phoenix dactylifera*) on experimentally-induced atypical prostatic hyperplasia in rats. J Inflamm 8(1): 1–13

El-Far AH, Shaheen HM, Abdel-Daim MM, Al Jaouni SK, Mousa SA (2016) Date palm (*Phoenix dactylifera*): protection and remedy food. Curr Trends Nutraceuticals 1(2):1–10

El-Far AH, Oyinloye BE, Sepehrimanesh M, Allah MAG, Abu-Reidah I, Shaheen HM, Razeghian-Jahromi I, Alsenosy AA, Noreldin AE, Al Jaouni SK, Mousa SA (2019) Date palm (*Phoenix dactylifera*): novel findings and future directions for food and drug discovery. Current Drug Discov Technol 16(1):2–10

Elhoumaizi MA, Panchanoor SD, Fang J, Chao CCT (2006) Confirmation of 'Medjool' date as a landrace variety through genetic analysis of Medjool accessions in Morocco. J Am Soc Hortic Sci 131:403–407

El-Houmaizi MA, Saadi M, Oihabi A, Cilas C (2002) Phenotypic diversity of date-palm cultivars (*Phoenix dactylifera* L.) from Morocco. Genet Resour Crop Evol 49:483–490

El Hadrami A, Daayf F, El Hadrami I (2011) Secondary metabolites of date palm. In: S. M. Jain, J. M. Al-Khayri and D.V. Johnson(Eds.). pp. 653–674. Date Palm Biotechnology, Springer, Netherlands.

Elhoumaizi MA, Saaidi M, Oihabi A, Cilas C (2002) Phenotypic diversity of date-palm cultivars (Phoenix dactylifera L.) from Morocco. Genetic resources and crop evolution 49(5):483–490

El-Kareem MG, Aal AA, Mohamed AY (2014) The synergistic effects of using silicon and selenium on fruiting of Zaghloul date palm (*Phoenix dactylifera* L.). Int J Agric Biosystems Eng 8(3):259–262

El-khishin DA, Adawy SS, Hussein EHA, Elitriby HA (2003) AFLP fingerprinting of some Egyptian date palm (*Phoenix dactylifera* L.) cultivars. Arab J Biotechnol 6:223–234

El-Kholy WM, Soliman TN, Darwish AMG (2019) Evaluation of date palm pollen (*Phoenix dactylifera* L.) encapsulation, impact on the nutritional and functional properties of fortified yoghurt. PLoS One 14(10):e0222789

Elmeer K, Mattat I (2012) Marker-assisted sex differentiation in date palm using simple sequence repeats. 3 Biotech 2(3):241–247

Elmeer K, Mattat I (2015) Genetic diversity of Qatari date palm using SSR markers. Genet Mol Res 14:1624–1635

Elsafy M, Zborowska A, Bryngelsson T et al (2016) Elucidating the genetic diversity of farmer cultivars of female date palms (*Phoenix dactylifera* L.) from Sudan by microsatellite markers. Genet Resour Crop Evol 63(6):975–986

Elshibli S (2009) Genetic diversity and adaptation of date palm (*Phoenix dactylifera* L.). Phd thesis, University of Helsinki, Helsinki, Finland, p 177

Elshibli S, Korpelainen H (2009) Excess heterozygosity and scarce genetic differentiation in the populations of *Phoenix dactylifera* L.: human impact or ecological determinants. Plant Genet Res 7:95–104. 2 Biodiversity, Genetic Diversity, and Genetic Resources of Date Palm 68

Elshibli S, Korpelainen H (2010) Identity of date palm (*Phoenix dactylifera* L.) germplasm in Sudan: from the morphology and chemical characters to molecular markers. Acta Hortic 859: 143–153

El-Shurafa MY, Ahmed HS, Abou-Naji SE (1982) Organic and inorganic constituents of date palm pit (in Libya). Date Palm J (FAO/NENADATES) eng 1(2):275–284

El-Sisy GA, El-Badry DA, El-Sheshtawy RI, El-Nattat WS (2018) Effects of *Phoenix dactylifera* pollen grains extract supplementation on post-thaw quality of Arabian stallion semen. Bulg J Vet Med 21(1):40–49

El-Sohaimy SA, Hafez EE (2010) Biochemical and nutritional characterizations of date palm fruits (*Phoenix dactylifera* L.). J Appl Sci Res 6(6):1060–1067

El-Tarras A, Hossni Y, lbanna EA, SefEldin A (2002) Genetic diversity within Egyptian date palm cultivars based on RAPD analysis. Egypt J Hortic 29:205–216

El-Yazal MA (2008) Sex determination of date palm (*Phoenix dactylifera* L.) through chemical composition of leaves. Fayoum J Agric Res Dev 22(2):76

Emoghene B, Asemota O, Idu M, Eke C (2015) Molecular characterization of some date palms in Nigeria using RAPD markers. J App Biol Biotech 3:26–30

Faisal HAI, Athafah QJ, Wahid AHA (2017) Effect of spraying iron chelated and salicylic acid on some characteristics of date palm (Phoenix dactylifera L.) Sayer cultivar. Kufa Journal for Agricultural Sciences 9(2):1–13

Faisal HA, Abdulla ASA (2018) The effect of foliar spray of potassium nitrate and chelated iron on some physiological and chemical characteristics for date palm Phoenix dactylifera L. Nersy cv. Basrah Journal for Date Palm Research 17(1–2):79–91

Fang Y, Wu H, Zhang T, Yang M, Yin Y, Pan L, Yu X, Zhang X, Hu S, Al- Mssalem IS, Yu J (2012) A complete sequence and transcriptomic analyses of date palm (*Phoenix dactylifera* L.) mitochondrial genome. PLoS One 7(5):e37164

Farhadi S, Ajerloo B, Mohammadi A (2017) Green biosynthesis of spherical silver nanoparticles by using date palm (*Phoenix dactylifera*) fruit extract and study of their antibacterial and catalytic activities. Acta Chim Slov 64(1):129–143

Fasal HA, Abdulwahid AH, Authafa QJ (2014) Effect of spray phosphorus and proline on some physical and chemical characteristics and production for date palm *Phoenix dactylifera* L. Shuker cultivar Hassan. Basrah J Date Palm Res 13(1–2):2–8

Fayadh JM, Al-Showiman SS (1990) Chemical composition of date palm (*Phoenix dactylifera* L.). J Chem Soc Pak 12(1):84–103

Fernández MI, Pedro JR, Seoane E (1983a) Two polyhydroxy stilbenes from stems of *Phoenix dactylifera*. Phytochemistry 22(12):2819–2821

Fernández MI, Pedro JR, Seoane E (1983b) Constituents of a hexane extract of *Phoenix dactylifera*. Phytochemistry 22(9):2087–2088

Frazer KA, Pachter L, Poliakov A et al (2004) VISTA: computational tools for comparative genomics. Nucleic Acids Res 1:32. (Web Server issue):W273–279

Gangwar AK, Ghosh AK, Saxena V (2014) Antiulcer activity of chloroform extract of *Phoenix dactylifera*Linn.Leaves. World J Pharm Res 3(2):3193–3200

Ghnimi S, Umer S, Karim A et al (2017) Date fruit (*Phoenix dactylifera* L.): an underutilized food seeking industrial valorization. NFS J 6:1–10

Ghosh T (2021) Date palm: An antimicrobial agent with its nutritional benefits. Int J Res Anal Scie Eng 1(2):14–14

Glasner B, Botes A, Zaid A, Emmens J (1999) Date harvesting, packing house management and marketing aspects. Date palm cultivation. FAO, Roma

González-Pérez MA, Caujapé-Castells J, Sosa PA (2004) Allozyme variation and structure of the Canarian endemic palm tree *Phoenix canariensis* (Arecaceae): implications for conservation. Heredity 93:307–315

González-Pérez MA, Caujapé-Castells J, Sosa PA (2004) Molecular evidence of hybridisation between the endemic Phoenix canariensis and the widespread P. dactylifera with Random Amplified Polymorphic DNA (RAPD) markers. Plant Systematics and Evolution 247:165–175

Gros-Balthazard M, Hazzouri KM, Flowers JM (2018) Genomic insights into date palm origins. Genes 9(10):502

Haas D (2015) Health and consciousness through ayurvedic cooking. MA Center, Kerala, India

Habashi AA, Kaviani M, Mousavi A, Khoshkam S (2008) Transient expression of b-glucuronidase reporter gene in date palm (*Phoenix dactylifera* L.) embryogenic calli and somatic embryos via microprojectile bombardment. J Food Agric Environ 6:160–163

Hachaichi A, Kouini B, Kian LK, Asim M, Jawaid M (2021) Extraction and characterization of microcrystalline cellulose from date palm fibers using successive chemical treatments. J Polym Environ 29(6):1990–1999

Hadrami AE, Daayf F, Hadrami IE (2011) Secondary Metabolites of Date Palm. In: Jain, S., Al-Khayri, J., Johnson, D. (eds) Date Palm Biotechnology. Springer, Dordrecht 653–674

Haider N, Nabulsi I, MirAli N (2012) Phylogenetic relationships among date palm (*Phoenix dactylifera* L.) cultivars in Syria using RAPD and ISSR markers. J Plant Biol Res 1:12–24

Haider MS, Khan IA, Jaskani MJ, Naqvi SA, Hameed M, Azam M, . . .& Pintaud JC (2015) Assessment of morphological attributes of date palm accessions of diverse agro-ecological origin. Pak J Bot 47(3):1143–1151

Hajian S, Hamidi-Esfahani Z (2015) Date palm status and perspective in Iran. In: Al-Khayri JM, Jain SM, Johnson DV (eds) Date palm genetic resources and utilization, vol 2: Asia and Europe. Springer, Dordrecht, pp 19–47

Hamad AM, Mustafa AI, Al-Kahtani MS (1983) Possibility of utilizing date syrup as sweetening and flavouring agent in ice cream making. In Proceedings of the first symposium on date palm in Saudi Arabia, King Faisal University, Al-Hassa, pp 544–550

Hamed AI, Ben Said R, Al-Ayed AS, Moldoch J, Mahael UA, Mahmoud AM, . . . Stochmal A (2017) Fingerprinting of strong spermatogenesis steroidal saponins in male flowers of *Phoenix dactylifera* (date palm) by LC-ESI-MS. Nat Prod Res 31(17):2024–2031

Hammadi H, Mokhtar R, Mokhtar E et al (2009) New approach for morphological identification of date palm (*Phoenix dactylifera* L.) cultivars from Tunisia. Pak J Bot 41:2671–2681

Hamwieh A, Farah J, Moussally S et al (2010) Development of 1000 microsatellite markers across the date palm (*Phoenix dactylifera* L.) genome. Acta Hortic 882:269–277

Hamza H, Benabderrahim MA, Elbekkay M, Ferdaous G, Triki T, Ferchichi A (2012) Investigation of genetic variation in Tunisian date palm (*Phoenix dactylifera* L.) cultivars using ISSR marker systems and their relation with fruit characteristics. Turk J Biol 36:449–458

Hanano A, Almousally I, Shaban M et al (2016) Biochemical, transcriptional, and bioinformatic analysis of lipid droplets from seeds of date palm (*Phoenix dactylifera* L.) and their use as potent

sequestration agents against the toxic pollutant, 2, 3, 7, 8- TetrachlorinatedDibenzo-p-dioxin. Front Plant Sci 7:836
Hassan A, Abou El Wafa MH (1947) An oestrogenic substance in pollen-grains of date palm tree *Phoenix dactylifera* L., Palmae. Nature 159(4038):409–410
Hattori K, Yamamoto Y, Fujii S, Kumrungsee T, Hasegawa M, Yoshida A, Suzuki T, Sambongi Y (2021) Fermented date residue extract mix containing gamma-aminobutyric acid augments the immune function of mouse splenocytes. Biosci Biotechnol Biochem
Hazzouri KM, Flowers JM, Visser HJ et al (2015) Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. Nat Commun 6:8824
Hazzouri KM, Gros-Balthazard M, Flowers JM et al (2019) Genome-wide association mapping of date palm fruit traits. Nat Commun 10(1):1–14
Hazzouri KM, Flowers JM, Nelson D et al (2020) Prospects for the study and improvement of abiotic stress tolerance in date palms in the post-genomics era. Front Plant Sci 11(293). https://doi.org/10.3389/fpls.2020.00293
He Z, Zhang C, Liu W et al (2017) DRDB: an online date palm genomic resource database. Front Plant Sci 8:1889
Heikrujam M, Sharma K, Prasad M et al (2015) Review on different mechanisms of sex determination and sex linked molecular markers in dioecious crops: a current update. Euphytica 201(2): 161–194
Henderson SA, Billotte N, Pintaud J-C (2006) Genetic isolation of Cape Verde Island *Phoenix atlantica* (Arecaceae) revealed by microsatellite markers. Conserv Genet 7:213–223
Hinkaew J, Aursalung A, Sahasakul Y, Tangsuphoom N, Suttisansanee U (2021) A comparison of the nutritional and biochemical quality of date palm fruits obtained using different planting techniques. Molecules 26(8):2245
https://www.harvestplus.org/where-we-work/india
https://www.mordorintelligence.com/industry-reports/date-market
Hussein F, El Zeid AA (1975) Chemical composition of Khalas dates grown in Saudi Arabia. Egyptian J Hortic (Egypt) 2:209–214.
Hyde J (1948) Processing Arizona dates. Arizona Date Institute, Phoenix. 20p
Ibraheem AEHA (2021) Food processing of date palm fruits in Sinai. In: Management and development of agricultural and natural resources in Egypt's desert. Springer, Cham, pp 491–515
Innih SO, Lorliam EB (2021) Anti-hyperlipidemic effect of date palm on hyperlipidemia in rats. J Anat Sci 12(1):117
Jaradat AA, Zaid A (2004) Quality traits of date palm fruits in a center of origin and center of diversity. Food Agric Environ 2:208–217
Jasim AM, Abbas MF, Shareef HJ (2016) Calcium application mitigates salt stress in date palm (*Phoenix dactylifera* L.) offshoots cultivars of Berhi and Sayer. Acta Agric Slov 107(1):103–112
Jindal VK, Mukherjee S (1969) Nature of sugars in Phoenix dactylifera seed polysaccharide. Cur Sci 459–460
Jindal VK, Mukherjee S (1970) Structure of alkali-soluble polysaccharide from *Phoenix dactilifera* seeds. Ind J Chem 8:417–419
Jubeir SM, Ahmed WA (2019) Effect of nanofertilizers and application methods on vegetative growth and yield of date palm. Iraqi J Agric Sci 50(1):267–274
Jubrael JMS (2001) Genetic characterization for some date palm cultivars in Iraq using RAPD markers. IPA J Agric Res 11:138–148. (in Arabic)
Jubrael JM, Udupa SM, Baum M (2005) Assessment of AFLP-based genetic relationships among date palm (*Phoenix dactylifera* L.) varieties of Iraq. J Americ Soc Hort Sci 130:442–447
Kabir MS, Akter MS, Hossain T, Khalil MI (2021) Microbiological quality assessment of date palm fruits and development of a decontamination method. Bangladesh J Bot 50(3):641–649
Kader AA, Hussein AM (2009) Harvesting and postharvest handling of dates, vol 4. ICARDA, Aleppo, p 15

Kader AA, Yahia EM (2011) Postharvest biology of tropical and subtropical fruits. In: Postharvest biology and technology of tropical and subtropical fruits. Woodhead Publishing, pp 79–111
Kadum H, Hamid AA, Abas F, Ramli NS, Mohammed AKS, Muhialdin BJ, Jaafar AH (2019) Bioactive compounds responsible for antioxidant activity of different varieties of date (*Phoenix dactylifera* L.) elucidated by 1H-NMR based metabolomics. Int J Food Prop 22(1):462–476
Kahramanoglu I, Usanmaz S (2019) Preharvest and postharvest treatments for increasing the rate of ripening of date palm fruit (*Phoenix dactylifera* L.) cv. Medjool. Prog Nutr 21(1):215–224
Karim K, Chokri B, Amel S, Wafa H, Richid H, Nouredine D (2010) Genetic diversity of Tunisian date palm germplasm using ISSR markers. Int J Bot 6:182–186
Khallouki F, Ricarte I, Breuer A, Owen RW (2018) Characterization of phenolic compounds in mature Moroccan Medjool date palm fruits (*Phoenix dactylifera*) by HPLC-DAD-ESI-MS. J Food Compos Anal 70:63–71
Khan A, Khan IA, Asif H et al (2010) Current trends in chloroplast genome research. Afr J Biotechnol 9(24):3494–3500
Khayyat M, Tafazoli E, Eshghi S, Rajaee S (2007) Effect of nitrogen, boron, potassium and zinc sprays on yield and fruit quality of date palm. Am Eurasian J Agric Environ Sci 2:289–296
Khierallah HSM (2007) Micropropagation of two date palm (*Phoenix dactylifera* L.) cultivars using inflorescences and study the genetic stability using AFLP-PCR markers. PhD dissertation, College of Agriculture, University of Baghdad, Bagdad, Iraq
Khierallah H, Bader SM, Baum M, Hamwieh A (2011) Assessment of genetic diversity for some Iraqi date palms (*Phoenix dactylifera* L.) using amplified fragment length polymorphisms (AFLP) markers. Afr J Biotechnol 10:9570–9576
Khierallah HSM, Al-Sammarraie SKI, Mohammed HI (2014) Molecular characterization of some Iraqi date palm cultivars using RAPD and ISSR markers. J Asia Sci Res 4:490–503
Kikuchi N, Miki T (1978) The separation of date (*Phoenix dactylifera*) sterols by liquid chromatography. Microchim Acta 69(1–2):89–96
Kim SG (2020) The way to true plant genome editing. Nat Plants 6(7):736–737
Kuhiyop EA, Adie DB, Abubakar UA (2020) Application of *Mangifera indica* (mango) and *Phoenix dactylifera* (dates) seeds powders as coagulants in wastewater treatment. Niger J Technol 39(1):269–277
Kumar N, Modi AR, Singh AS, Gajera BB, Patel AR, Patel MP, Subhash N (2010) Assessment of genetic fidelity of micropropagated date palm (*Phoenix dactylifera* L.) plants by RAPD and ISSR markers assay. Physiol Mol Biol Plants 16:207–213
Kumar S, Mukherjee A, Dutta J (2020) Chitosan based nanocomposite films and coatings: emerging antimicrobial food packaging alternatives. Trends Food Sci Technol 97:196–209
Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 29:4633–4642
Lamia FS, Mukti RF (2021) Bangladeshi wild date palm fruits (*Phoenix sylvestris*): promising source of anti-cancer agents for hepatocellular carcinoma treatment. Int J Appl Sci Biotechnol 9 (1):32–37
Lee DJ, Schoenberger R, Archibald J, McCollum S (2008) Development of a machine vision system for automatic date grading using digital reflective near-infrared imaging. J Food Eng 86 (3):388–398
Liolios CC, Sotiroudis GT, Chinou I (2009) Fatty acids, sterols, phenols and antioxidant activity of *Phoenix theophrasti* fruits growing in Crete, Greece. Plant Foods Hum Nutr 64(1):52–61
López-Córdoba A (2021) Technologies for the development of new value-added foods from dates and their processing by products. Preprints.org 2021, 2021050555. https://doi.org/10.20944/preprints202105.0555.v1.
Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25(5):955–964
Mahar AQ (2007) Post-harvest studies of different cultivars of date palm (*Phoenix dactylifera* L.) fruits, their protection, identification, processing and preservation at district Khairpur, Sindh,

Pakistan. Ph.D thesis, Date Palm Research Institute, Shah Abdul Latif University, Khairpur, Sindh, Pakistan
Mahran GH, Abdel Whahab SM, Attia AM (1976) A phytochemical study of date palm pollen. Planta Med 29(2):171–175. Pascal and Francis Bibliographic Databases
Makhlouf-Gafsi I, Mokni-Ghribi A, Bchir B et al (2016) Physico-chemical properties and amino acid profiles of sap from Tunisian date palm. Sci Agric 73:85–90
Malek JA, Mathew S, Mathew LS et al (2020) Deletion of beta-fructofuranosidase (invertase) genes is associated with sucrose content in date palm fruit. Plant Direct 4(5):e00214
Manju KP, Manimekalai R, Naganeeswaran SA et al (2016) Microsatellites mining in date palm (*Phoenix dactylifera* L.) and their cross transferability across Arecaceae family. Plant Omics 9 (3):191–197
Mao Y, Botella JR, Liu Y, Zhu JK (2019) Gene editing in plants: progress and challenges. Natl Sci Rev 6(3):421–437
Maou S, Meghezzi A, Grohens Y, Meftah Y, Kervoelen A, Magueresse A (2021) Effect of various chemical modifications of date palm fibers (DPFs) on the thermo-physical properties of polyvinyl chloride (PVC)–high-density polyethylene (HDPE) composites. Ind Crop Prod 171:113974
Markhand GS, Abul-Soad AA, Mirbahar AA, Kanahr NA (2010) Fruit characterization of Pakistani dates. Pak J Bot 42:3715–3722
Martínez Vega M, Varming C, Skov T, Toldam-Andersen TB (2014) Post-harvest ripening increase cultivar specific sensory and analytical aroma profile in apple juice: a study of four commercial cultivars in Denmark. Acta Agric Scand B Soil Plant Sci 64(3):244–251
Maryam JMJ, Awan FS, Ahmad S, Khan IA (2016) Development of molecular method for sex identification in date palm (*Phoenix dactylifera* L.) plantlets using novel sex-linked microsatellite markers. 3 Biotech 6:22
Mathew LS, Spannag M, Al-Malki A et al (2014) A first genetic map of date palm (*Phoenix dactylifera*) reveals long-range genome structure conservation in the palms. BMC Genomics 15:285
Mathew LS, Seidel MA, George B, Mathew S, Spannagl M, Haberer G, Torres MF, Al-Dous EK, Al-Azwani EK, Diboun I, Krueger RR, Mayer KFX, Mohamoud YA, Suhre K, Malek JA (2015) A genome-wide survey of date palm cultivars supports two major subpopulations in *Phoenix dactylifera*. G3 (Bethesda) 5:1429–1438
Megbo BC, Samuel AM, Dio DW (2017) *Phoenix dactylifera* fruit: a nutraceutical agent in the treatment of diarrhea. Innovat Int J Med Pharm Sci 2(3). https://doi.org/10.24018/10.24018/iijmps.2018.v1i1.22
Mirbahar AA, Markhand GS, Khan S, Abul-Soad AA (2013) Genetic diversity analysis among important date palm (*Phoenix dactylifera* L.) cultivars from Pakistan using ISSR markers. In: Abstracts of international conference on date palm: present status &future prospects, The Islamia University of Bahawalpur, Bahawalpur, 2–3 Sept, p 26
Mirbahar AA, Markhand GS, Khan S, Abul-Soad AA (2014) Molecular characterization of Pakistani date palm (*Phoenix dactylifera* L.) cultivars. Pak J Bot 46:619–625
Moghaieb REA, Abdel-Hadi AA, Ahmed MRA, Hassan AGM (2010) Genetic diversity and sex determination in date palms (*Phoenix dactylifera* L.) based on DNA markers. Arab J Biotech 13: 143–156
Mohammadi M, Soltani M, Siahpoosh A, Hosseini Shekarabi SP, ShamsaieMehrgan M, Lymbery A (2018) Effect of date palm (*Phoenix dactylifera*) seed extract as a dietary supplementation on growth performance immunological haematological biochemical parameters of common carp. Aquac Res 49(8):2903–2912
Mohamoud YA, Mathew LS, Torres MF et al (2019) Novel subpopulations in date palm (*Phoenix dactylifera*) identified by population-wide organellar genome sequencing. BMC Genomics 20 (1):498
Mohei EL, Mohasseb HAA, Al-Khateeb AA et al (2019) Identification and sequencing of date-SRY gene: a novel tool for sex determination of date palm (*Phoenix dactylifera* L.). Saudi J Biol Sci 26(3):514–523

Mokhtar MM, Adawy SS, El-Assal SEDS et al (2016) Genic and intergenic SSR database generation, SNPs determination and pathway annotations, in date palm (*Phoenix dactylifera* L.). PLoS One 11(7):0159268

Moslemi E, Dehghan P, Khani M (2022) The effect of date seed (*Phoenix dactylifera*) supplementation on inflammation, oxidative stress biomarkers, and performance in active people: a blinded randomized controlled trial protocol. Contemp Clin Trials Commun 28:100951

Mossa JS, Hifnawy MS, Mekkawi AG (1986) Phytochemical and biological investigations on date seeds (*Phoenix dactylifera* L.) produced in Saudi Arabia. Arab Gulf J Sci Res

Mousavi M, Mousavi A, Habashi AA, Dehsara B (2014) Genetic transformation of date palm (*Phoenix dactylifera* L. cv.'Estamaran') via particle bombardment. Mol Biol Rep 41(12):8185–8194

Mueller UG, Wolfenbarger LL (1999) AFLP genotyping and fingerprinting. Trends Ecol Evol 14: 389–394

Munshi A, Osman G (2010) Investigation on molecular phylogeny of some date palm (*Phoenix dactylifera* L.) cultivars by protein, RAPD and ISSR markers in Saudi Arabia. Austr J Crop Sci 4:23–28

Nadeem M, Qureshi TM, Ugulu I, Riaz MN, An QU, Khan ZI, ... Dogan Y (2019) Mineral, vitamin and phenolic contents and sugar profiles of some prominent date palm (*Phoenix dactylifera*) varieties of Pakistan. Pak J Bot 51(1):171–178

Naeem A, Khan SH, Khan IA, Khan AA (2018) SSR-based genetic diversity of date palm in Makran (Pakistan). Pak J Agric Sci 55(4):863

Nehdi IA, Sbihi HM, Tan CP, Rashid U, Al-Resayes SI (2018) Chemical composition of date palm (*Phoenix dactylifera* L.) seed oil from six Saudi Arabian cultivars. J Food Sci 83(3):624–630

Njoroge JB, Ninomiya K, Kondo N, Toita H (2002) Automated fruit grading system using image processing. In: Proceedings of the 41st SICE annual conference.SICE 2002, vol 2. IEEE, pp 1346–1351

Noorbakhsh H, Khorasgani MR (2022) Date (*Phoenix dactylifera* L.) polysaccharides: a review on chemical structure and nutritional properties. J Food Meas Charact 21:1–1

Nwanekezi EC, Ekwe CC, Agbugba RU (2015) Effect of substitution of sucrose with date palm (*Phoenix dactylifera*) fruit on quality of bread. J Food Process Technol 6(9):1

Nyika JM (2022) Nanotechnology and its applications in environmental remediation. In: Tripathi S, Sanjeevi R, Anuradha J, Chauhan DS, Rathoure AK (eds) Nano-bioremediation: nanotechnology and bioremediation. In research anthology on emerging techniques in environmental remediation. IGI Global, pp 135–149

Onyilo PO, Innih SO, Ihemeson G, Ogereka LU, Nweke SM, Obele EI, Nwalie IN (2021) Ameliorative effects of aqueous extract of date palm fruits (*Phoenix dactylifera*) on high fat diet-induced liver damage in adult wistar rats. J Appl Sci Environ Manag 25(3):327–331

Osman AMA (1984) The performance of date palms in The Sudan. Acta Hortic 143:231–237

Otify AM, El-Sayed AM, Michel CG, Farag MA (2019) Metabolite profiling of date palm (*Phoenix dactylifera* L.) commercial by-products (pits and pollen) in relation to its antioxidant effect: a multiplex approach of MS and NMR metabolomics. Metabolomics 15(9):1–17

Otify AM, Hammam AMM, Farag MA (2021) *Phoenix dactylifera* L. date tree pollen fertility effects on female rats in relation to its UPLC-MS profile via a biochemometric approach. Steroids 173:108888

Ould Mohamed Salem A, Rhouma S, Zehdi S et al (2008) Morphological variability of Mauritanian date palm (*Phoenix dactylifera* L.) cultivars as revealed by vegetative traits. Acta Bot Croatica 67:81–90

Pandey SN, Ghosh SK (1995) The chemical nature of date-palm (*Phoenix dactylifera*-l) leaf fibre. J Text Inst:487–489

Patankar HV, Al-Harrasi I, Al-Yahyai R et al (2019) Functional characterization of date palm aquaporin gene PdPIP1; 2 confers drought and salinity tolerance to yeast and Arabidopsis. Genes-Basel 10(5):390

Pathak MR, Hamzah R (2008) RAPD analysis of date palm cultivars of Bahrain. FloriOrn Biotech 2:9–11

Premkrishnan BV, Arunachalam V (2012) Insilico RAPD priming sites in expressed sequences and iSCAR markers for oil palm. Int J Genom 2012. https://doi.org/10.1155/2012/913709
Qacif N, Baaziz M, Bendiab K (2007) Biochemical investigations on peroxidase contents of male and female inflorescences of date palm (*Phoenix dactylifera* L.). Sci Hortic 114:298–301
Qiao K, Wang F, Liang S, Wang H, Hu Z, Chai T (2019) New biofortification tool: wheat TaCNR5 enhances zinc and manganese tolerance and increases zinc and manganese accumulation in rice grains. J Agric Food Chem 67(35):9877–9884
Rabei S, Said WM, Rizk RM, El-Sharabasy SF (2012) Morphometric taxonomy of date palm diversity growing in Egypt. In: Egypt. J Bot, 2nd international conference, Minia University, Egypt, 29–30 Apr, pp 175–189
Racchi ML, Bove A, Turchi A, Bashir G, Battaglia M, Camussi A (2014) Genetic characterization of Libyan date palm resources by microsatellite markers. 3 Biotech 4(1):21–32
Rahmawati M, Mahfud C, Risuleo G, Jadid N (2022) Nanotechnology in plant metabolite improvement and in animal welfare. Appl Sci 12(2):838
Rajan M, Shanmugam S, Guedes TJFL, Thangaraj P, Barbosa PF, Narain N (2021) Profiles of nutritional, bioactive compounds and cytotoxic activity of dwarf date palm (*Phoenix loureiroi*) fruit. J Food Meas Charact 15(2):1611–1625
Rani R, Kharb P, Uppal S, Dhillon S, Chowdhury VK (2007) Molecular characterization and genetic divergence analysis in date palm (*Phoenix dactylifera* L.) using RAPD markers. Haryana J Hort Sci 36(12):101–103
Rao PVKJ, Das M, Das SK (2009) Changes in physical and thermo-physical properties of sugarcane, Palmyra palm and date-palm juices at different concentration of sugar. J Food Eng 90:559–566
Rasheed S, Zaidi S, Azim MK (2020) The chloroplast genome sequence of *Momordica charantia* L. (bitter gourd). Gene Rep 21:100963
Refaai MM (2014) Response of Zaghloul date palms grown under Minia region conditions to spraying wheat seed sprout extract and nano-boron. Stem Cells 5(4):22–28
Ressan SH, Al-Tememi HH (2019) Study of evaluating the efficiency of biological and Nano-fertilizers treatments and their addition methods in some the physiological and productive characteristics of date palm Phoenix dactlifera L. Basrah Journal for Date Palm Research 18(2):84–101
Rivera D, Obón de Castro C, Carreno E et al (2008) Morphological systematic of date palm diversity (*Phoenix*, Arecaceae) in Western Europe and some preliminary molecular results. Acta Hortic 799:97–104
Rose DC, Chilvers J (2018) Agriculture 4.0: broadening responsible innovation in an era of smart farming. Front Sustain Food Syst 2:87
Saafi EB, Trigui M, Thabet R, Hammami M, Achour L (2008) Common date palm in Tunisia: chemical composition of pulp and pits. Int J Food Sci Technol 43(11):2033–2037
Sabir JSM, Abo-Aba S, Bafeel S, Zari TA, Edris S, Shokry AM, Atef A, Gadalla NO, Ramadan AM, Al-Kordy MA, El-Domyati FM, Jansen RK, Bahieldin A (2014a) Characterization of ten date palm (*Phoenix dactylifera* L.) cultivars from Saudi Arabia using AFLP and ISSR markers. C R Biol 337:6–18
Sabir JSM, Arasappan D, Bahieldin A, Abo-Aba S, Bafeel S, Zari TA, Edris S, Shokry AM, Gadalla NO, Ramadan AM, Atef A, Al-Kordy MA, El-Domyati FM, Jansen RK (2014b) Whole mitochondrial and plastid genome SNP analysis of nine date palm cultivars reveals plastid heteroplasmy and close phylogenetic relationships among cultivars. PLoS One 9:1–13
Sackey J, Akbari M, Morad R, Bashir AKH, Ndiaye NM, Matinise N, Maaza M (2021) Molecular dynamics and biosynthesis of *Phoenix dactylifera* mediated Mn_3O_4 nanoparticles: electrochemical application. J Alloys Compd 854:156987
Safronov O, Kreuzwieser J, Haberer G et al (2017) Detecting early signs of heat and drought stress in *Phoenix dactylifera* (date palm). PLoS One 12(6):0177883
Saker MM, Adawy SS, Mohamed AA, El-Itriby HA (2006) Monitoring of cultivar identity in tissue culture-derived date palms using RAPD and AFLP analysis. Biol Plant 50:198–204

Saker M, Allam MA, Goma AH et al (2007) Optimization of some factors affecting genetic transformation of semi-dry Egyptian date palm cultivar (Sewi) using particle bombardment. J Genet Eng Biotechnol 5:1–6

Saker M, Ghareeb H, Kumlehn J (2009) Factors influencing transient expression of agrobacterium mediated transformation of GUS gene in embryogenic callus of date palm. Adv Hortic Sci 23: 150–157

Sakka H, Zehdi S, Ould MSA, Rhouma A, Marrakchi M, Trifi M (2003) Tunisian date palm (*Phoenix dactylifera* L.) genotypes identification mediated by plastid PCR/RFLP based DNA. J Genet Breed 57:259–264

Salih AM, Al-Qurainy F, Khan S, Tarroum M, Nadeem M, Shaikhaldein HO, Gaafar AZ, Alfarraj NS (2021) Biosynthesis of zinc oxide nanoparticles using *Phoenix dactylifera* and their effect on biomass and phytochemical compounds in Juniperus procera. Sci Rep 11(1):1–12

Sarraf M, Jemni M, Kahramanoğlu I, Artés F, Shahkoomahally S, Namsi A, Ihtisham M, Brestic M, Mohammadi M, Rastogi A (2021) Commercial techniques for preserving date palm (*Phoenix dactylifera*) fruit quality and safety: a review. Saudi J Biol Sci 28(8):4408

Sattar MN, Iqbal Z, Tahir MN, Shahid MS, Khurshid M, Al-Khateeb AA, Al-Khateeb SA (2017) CRISPR/Cas9: a practical approach in date palm genome editing. Front Plant Sci 8:1469

Sawaya WN, Khatchadourian HA, Khalil JK, Safi WM, Al-Shalhat A (1982) Growth and compositional changes during the various developmental stages of some Saudi Arabian date cultivars. J Food Sci 47(5):1489–1492

Sawaya WN, Khalil JK, Safi WJ, Khatchadourian HA (1983a) Date bars fortified with soy protein isolate and dry skim milk. J Food Sci 48(5):1503–1506

Sawaya WN, Safi WM, Khalil JK, Khatchadourian HA, Moshadi AS (1983b) Sugars, Tannins and some vitamins content of twenty-Five date cultivars grown in Saudi Arabia at the Khalal (Mature colure) and Tamre (Rip) stages. First symposium on date palm King Faisal, Al-Hassa-Saudi Arabia, pp 468–478. Canadian Inst

Sedra MH (2013) Genetic diversity analysis of Moroccan cultivar genotyping and rapid screening for Bayoud disease resistance in date palm using molecular techniques. Acta Hortic 994: 271–286

Sedra MH (2015) Date palm status and perspective in Morocco. In: Al-Khayri JM, Jain SM, Johnson DV (eds) Date palm genetic resources and utilization, vol 1. Africa and the Americas. Springer, Dordrecht, pp 257–323

Sedra M, Lazrek BH (2011) Fusarium oxysporum f. sp. albedinis toxin characterization and use for selection of resistant date palm to bayoud disease. Date Palm Biotechnology, 253–270. https://www.statista.com/statistics/960213/date-palmmarket-value-worldwide/

Sedra MH, Lashermes P, Trouslot M, Hamon S (1998) Identification and genetic diversity analysis of date palm (*Phoenix dactylifera* L.) varieties of Morocco using RAPD markers. Euphytica 103:75–82

Sellappan L, Manoharan S, Sanmugam A, Anh NT (2022) Role of nanobiosensors and biosensors for plant virus detection. In: Nanosensors for smart agriculture. Elsevier, pp 493–506

Sghaier B, Kriaa W, Bahloul M, Novo JVJ, Drira N (2009) Effect of ABA, arginine and sucrose on protein content of date palm somatic embryos. Sci Hortic 120(3):379–385

Shaba EY, Ndamitso MM, Mathew JT, Etsunyakpa MB, Tsado AN, Muhammad SS (2015) Nutritional and anti-nutritional composition of date palm (*Phoenix dactylifera* L.) fruits sold in major markets of Minna Niger state, Nigeria. Afr J Pure Appl Chem 9(8):167–174

Shahbaz K, Asif JA, Liszen T, Nurul AA, Alam MK (2022) Cytotoxic and antioxidant effects of *Phoenix dactylifera* L.(Ajwa date extract) on oral squamous cell carcinoma cell line. Biomed Res Int 2022:5792830

Sham A, Aly MA (2012) Bioinformatics based comparative analysis of omega-3 fatty acids in desert plants and their role in stress resistance and tolerance. Int J Plant Res 2:1

Shareef HJ (2011a) Effect of spraying with urea and NPK on production of date palm *Phoenix dactylifera* L. cv. Khidrawi. Basrah J Date Palm Res 10(1):56–67

Shareef HJ (2011b) Effect of spraying with urea and NPK on some chemical characteristics and ripening of date palm fruits *Phoenix dactylifera* L. cv. Khidrawi. Jo Basrah Res (Sciences) 37(4B):3–25

Shareef HJ (2019) Salicylic acid and potassium nitrate promote flowering through modulating the hormonal levels and protein pattern of date palm *Phoenix dactylifera* 'Sayer' offshoot. Acta Agric Slov 114(2):231–238

Shareef HJ, Al-Yahyai RA, Omar AEDK, Barus WA (2020) Foliar nano-fertilization enhances fruit growth, maturity, and biochemical responses of date palm. Can J Plant Sci 101(3):299–306

Sharifi M, Sheidai M, Koohdar F (2018) Genetic fingerprinting of date palm (*Phoenix dactylifera* L.) by using ISSR and cpDNA sequences. Indian J Genet Plant Breed 78(04):507–514

Sharma M, Chai C, Morohashi K, Grotewold E, Snook ME, Chopra S (2012) Expression of flavonoid 3′-hydroxylase is controlled by P1, the regulator of 3-deoxyflavonoid biosynthesis in maize. BMC Plant Biol 12(1):1–13

Shehzad M, Rasheed H, Naqvi SA, Al-Khayri JM, Lorenzo JM, Alaghbari MA, Manzoor MF, Aadil RM (2021) Therapeutic potential of date palm against human infertility: a review. Metabolites 11(6):408

Shockey J (2020) Gene editing in plants: assessing the variables through a simplified case study. Plant Mol Biol 103:75–89.

Siddiqi SA, Rahman S, Khan MM, Rafiq S, Inayat A, Khurram MS et al (2020) Potential of dates (*Phoenix dactylifera* L.) as natural antioxidant source and functional food for healthy diet. Sci Total Environ 748:141234

Silabdi S, Khali M, Tenore GC, Stiuso P, Vanacore D, Novellino E (2021) *Phoenix dactylifera* polyphenols improve plasma lipid profile in hyperlipidemic rats and oxidative stress on HepG2 cells. Int J Geogr Inf Syst 27(2):161–176

Singh P, Kharb P, Pandey N, Batra P, Khatak S, Dhillon S, Uppal S, Chowdhury VK (2006) RAPD analysis for genetic diversity and identification of sex-specific marker in date palm (*Phoenix dactylifera* L.). Haryana J Hort Sci 35:232–234

Singh D, Rajawat MVS, Kaushik R, Prasanna R, Saxena AK (2017) Beneficial role of endophytes in biofortification of Zn in wheat genotypes varying in nutrient use efficiency grown in soils sufficient and deficient in Zn. Plant Soil 416(1):107–116

Sirry SM, Ali S, Abdelaziz A, Mohamed A (2020) Biosynthesis of silver nanoparticles using the extracts of palm date seeds (*Phoenix dactylifera* L.) and its application in increasing the anti-inflammatory effect of piroxicam drug. Adv Nat Sci Nanosci Nanotechnol 11(3):035017

Sivarajasekar N, Prakashmaran J, Naushad M et al (2019) Recent updates on heavy metal remediation using date stones (Phoenix dactylifera L.) – date fruit processing industry waste. In: Naushad M, Lichtfouse E (eds) Sustainable agriculture reviews, vol 34. Springer, Cham, pp 193–206

Soliman S, Bahy A, Mohamed Morsy M (2003) Genetic comparisons of Egyptian date palm cultivars (*Phoenix dactylifera* L.) by RAPD-PCR. Afr J Biotechnol 2:86–87

Soliman KH, Rizk RM, El-Sharabasy SF (2006) Genetic polymorphism of semi-dry date palm (*Phoenix dactylifera* L.) cultivars in Egypt. Aust J Biotechnol 22:261–273

Soliman SS, Al-Obeed RS, Omar AA et al (2013) A comparative study of the morphological characteristics of some seedling date palm males. J Appl Sci Res 9(7):4463–4468

Soliman SS, Al-Obeed RS, Ahmed TA (2015) Physico-chemical characteristics of oil produced from seeds of some date palm cultivars (*Phoenix dactylifera* L). J Environ Biol 36(2):455

Sonia DDS, Laurent CL, Cosette P et al (2013) The date palm (*Phoenix dactylifera* L.) leaf proteome: identification of a gender biomarker to screen male parents. Pak Orthodontic J 6 (1):18–23

Soumaya RC, Sarra C, Salwa ZA, Khaled C, Khaled S (2014) Molecular polymorphism and phylogenetic relationships within Tunisian date palm (Phoenix dactylifera L.): evidence of noncoding trnL-trnF regions of chloroplast DNAs. Scientia Horticulturae 170:32–38

Srivastav VS, Kapadia CV, Mahatma MK, Jha SK, Jha S, Ahmad T (2013) Genetic diversity analysis of date palm (*Phoenix dactylifera* L.) in the Kutch region of India using RAPD and ISSR markers. Emir J Food Agric 25:907–915

Stangoulis JC, Knez M (2022) Biofortification of major crop plants with iron and zinc-achievements and future directions. Plant Soil 17:1–20

Sudhersan C, Al-Sabah L, Jibi Manuel S (2010) Xenic and metaxenic effect of Phoenix pusilla pollen on certain date palm cultivars. In IV International Date Palm Conference 882, pp. 297–302

Sulieman AME, Itimad A, Abd-Elhafise Abdelrahim AM (2012) Comparative study on five Sudanese date (*Phoenix dactylifera* L.) fruit cultivars. Food Nutr Sci 3:1245–1251

Suzuki K, Konno R, Shimizu T, Nagashima T, Kimura A (2007) A fermentation product of phytosterol including campestenone reduces body fat storage and body weight gain in mice. J Nutr Sci Vitaminol 53(1):63–67

Taha FH, Abood MR (2018) Influence of some organic fertilizers on date palm cv. Barhi. Iraqi J Agric Sci 49(4):632–638

Takdehghan G, Dorchin M, Amin M, Ghorbanzadeh B, Assarian AR, Amirgholami N, Amin G (2021) Efficacy of date palm (*Phoenix dactylifera* L.) leaf extract in COVID-19 infection: a randomized double-blind clinical trial. MAR Oncol, 1(1).

Takele D, Tsegaw M, Indracanti M (2021) Genetic diversity assessment in some landraces and varieties of date palm (*Phoenix dactylifera* L.) from Afar region, Ethiopia, using ISSR markers. Ecol Genet Genom 19:100085

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Tawfek MA, Baker EA, El-Sayed HA (2021) Study properties of fermented camels' and goats' milk beverages fortified with date palm (*Phoenix dactylifera* L.). Food Nutr Sci 12(5):418–428

Tengberg M (2012) Beginning and early history of date palm garden cultivation in the Middle East. J Arid Environ 86:139–147

Terral JF, Newton C, Ivorra S et al (2012) Insights into the historical biogeography of the date palm (*Phoenix dactylifera* L.) using geometric morphometry of modern and ancient seeds. J Biogeogr 39(5):929–941

Thareja G, Mathew S, Mathew LS et al (2018) Genotyping-by-sequencing identifies date palm clone preference in agronomics of the State of Qatar. PLoS One 13(12):0207299

Thiel T, Michalek W, Varshney R et al (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106(3):411–422

Toghueo RMK (2019) Anti-leishmanial and anti-inflammatory agents from endophytes: a review. Nat Prod Bioprospecting 9(5):311–328

Tong J, Sun M, Wang Y, Zhang Y, Rasheed A, Li M, Xia X, He Z, Hao Y (2020) Dissection of molecular processes and genetic architecture underlying iron and zinc homeostasis for biofortification: from model plants to common wheat. Int J Mol Sci 21(23):9280

Toor AK, Sharma R, Singh G, Mohaptra T (2005) Genetic variation in date palm (*Phoenix dactylifera* L.) cultivars assessed using RAPD markers. Indian J Genet Plant Breed 65(4): 335–336

Torres MF, Mathew LS, Ahmed I, Al-Azwani IK, Krueger R, Rivera-Nuñez D, Mohamoud YA, Clark AG, Suhre K, Malek JA (2018) Genus-wide sequencing supports a two-locus model for sex-determination in Phoenix. Nature communications 9(1):3969

Trifi M, Rhouma A, Marrakchi M (2000) Phylogenetic relationships in Tunisian date palm (*Phoenix dactylifera* L.) germplasm collection using DNA amplification fingerprinting. Agronomie 20: 665–671

Triki MA, Zouba A, Khoualdia O, Mahamoud OB, Takrouni ML, Garnier M, Bové JM, Montarone M, Poupet A, Flores R, Darós JA, Fadda ZGN, Moreno P, Durán-Vila N (2003) "Maladie des feuilles cassantes" or brittle leaf disease of date palms in Tunisia: Biotic or abiotic disease? Journal of Plant Pathology 71–79

Tsanova T, Stefanova L, Topalova L, Atanasov A, Pantchev I (2021) DNA-free gene editing in plants: a brief overview. Biotechnol Biotechnol Equip 35(1):131–138

Uddin MS, Nuri ZN (2021) Nutritional values and pharmacological importance of date fruit (*Phoenix dactylifera* Linn): a review. J Curr Res Food Sci 2(1):27–30
Uhl NW, Dransfield J (1987) Genera palmarum: a classification of palms based on the work of Harold E. Moore, Jr, L.H. baily. Hortorium and the International Palm Society, Lawrence, pp 214–217
Vats S, Kumawat S, Kumar V, Patil GB, Joshi T, Sonah H, Sharma TR, Deshmukh R (2019) Genome editing in plants: exploration of technological advancements and challenges. Cell 8 (11):1386
Vayalil PK (2002) Antioxidant and antimutagenic properties of aqueous extract of date fruit (*Phoenix dactylifera* L. Arecaceae). J Agric Food Chem 50(3):610–617
Viana VE, Maltzahn LE, Costa de Oliveira A, Pegoraro C (2021) Genetic approaches for iron and zinc biofortification and arsenic decrease in *Oryza sativa* L. grains. Biol Trace Elem Res: 200(10):4505-4523. https://doi.org/10.1007/s12011-021-03018-0.
Walsh KB, Blasco J, Zude-Sasse M, Sun X (2020) Visible-NIR 'point' spectroscopy in postharvest fruit and vegetable assessment: the science behind three decades of commercial use. Postharvest Biol Technol 168:111246
Wang Y, Chao X, Shi H, Mehboob H, Hassan W (2019) *Phoenix dactylifera* protects against doxorubicin-induced cardiotoxicity and nephrotoxicity. Cardiol Res Pract 2019:7395239
Wyman S (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20 (17):3252–3255
Xiao Y, Xia W, Ma J et al (2016) Genome-wide identification and transferability of microsatellite markers between Palmae species. Front Plant Sci 7:1578
Xin C, Liu W, Lin Q, Zhang X, Cui P, Li F et al (2015) Profiling microRNA expression during multi-staged date palm (*Phoenix dactylifera* L.) fruit development. Genomics 105:242–251
Yahia EM, Lobo MG, Kader AA (2014) Harvesting and postharvest technology of dates. In: Siddiq M, Aleid SM, Kader AA (eds) Dates: postharvest science, processing technology and health benefits, 1st edn, Wiley-Blackwell, pp 105–135
Yaish MW, Sunkar R, Zheng Y, Ji B, Al-Yahyai R, Farooq SA (2015) A genome-wide identification of the miRNAome in response to salinity stress in date palm (*Phoenix dactylifera* L.). Front Plant Sci 6:946
Yaish MW, Patankar HV, Assaha DV, Zheng Y, Al-Yahyai R, Sunkar R (2017) Genome-wide expression profiling in leaves and roots of date palm (*Phoenix dactylifera* L.) exposed to salinity. BMC Genomics 18(1):1–17
Yang M, Zhang X, Liu G, Yin Y, Chen K, Yun Q, Zhao D, Al-Mssalem IS, Yu J (2010) The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). PLoS One 5(9): e12762
Younis RA, Ismail OM, Soliman SS (2008) Identification of sex-specific DNA markers for date palm (*Phoenix dactylifera* L.) using RAPD and ISSR techniques. Res J Agric Biol Sci 4(4):278–284
Yusuf AO, Culham A, Aljuhani W et al (2015) Genetic diversity of Nigerian date palm (Phoenix dactylifera) germplasm based on microsatellite markers. Int J Biosci Biotechnol 7(1):121–132
Zabar AF, Borowy A (2012) Cultivation of date palm in Iraq. Ann Univ Mariae Curie Sklodowska 22:39–54
Zafar S, Zafar A (2019) Biosynthesis and characterization of silver nanoparticles using *Phoenix dactylifera* fruits extract and their in vitro antimicrobial and cytotoxic effects. Open Biotechnol J 13(1):37–46
Zaid A, De Wet PF (2002a) Climatic requirements of date palm. In: Zaid, A. and Arias-Jiménez, E.J. Date palm cultivation. Food andAgricultural Organization of the United Nations, FAO Plant production andprotection papers, 156 Rev.1. Rome pp 57–72
Zaid A, de Wet PF (2002b) Pollination and bunch management. In: Zaid A (ed) Date palm cultivation. FAO plant production & proection aper 156 rev, vol 1, pp 227–269
Zain M, Yasmeen H, Yadav SS, Amir S, Bilal M, Shahid A, Khurshid M (2022) Applications of nanotechnology in biological systems and medicine. In: Nanotechnology for Hematology, blood transfusion, and artificial blood. Elsevier, pp 215–235

Zayneb C, Imen RH, Walid K et al (2017) The phytochelatin synthase gene in date palm (*Phoenix dactylifera* L.): phylogeny, evolution and expression. Ecotoxicol Environ Saf 140:7–17

Zehdi S, Trifi M, Salem AOM, Rhouma A, Marrakchi M (2002) Survey of inter simple sequence repeat polymorphisms in Tunisian date palms (*Phoenix dactylifera* L.). J Genet Breed 56:77–83

Zehdi S, Sakka H, Rhouma A, Salem AOM, Marrakchi M, Trifi M (2004a) Analysis of Tunisian date palm using simple sequence repeat primers. Afr J Biotechnol 3:215–219

Zehdi S, Trifi M, Billotte N, Marrakchi M, Pintaud JC (2004b) Genetic diversity of Tunisian date palms (*Phoenix dactylifera* L.) revealed by nuclear microsatellite polymorphism. Hereditas 141: 278–287

Zehdi S, Sakka H, Salem AOM, Rhouma A, Marrakchi M, Trifi M (2005) Molecular polymorphism and genetic relationships in a Tunisian date palm (*Phoenix dactylifera* L.) collection using ISSR amplification finger printing. Plant Genet Resour Newsl 144:39–44

Zehdi S, Pintaud JC, Billotte N, Salem AOM, Sakka H, Rhouma A, Marrakchi M, Trifi M (2006) Use of SSR markers to provide an identification key for date palm varieties (*Phoenix dactylifera* L.). Plant Genet Resour Newsl 145:11–18

Zehdi-Azouzi S, Rhouma S, Dkkhil-Dakhlaoui S, Salem AOM, Cherif E, Othmani A, Marrakchi M, Trifi M (2011) Polymorphism and genetic relationship in date palm using molecular markers. In: Jain SM, Al-Khayri JM, Johnson DV (eds) Date palm biotechnology. Springer, Dordrecht, pp 407–425

Zhang G, Pan L, Yin Y et al (2012) Large-scale collection and annotation of gene models for date palm (*Phoenix dactylifera*, L.). Plant Mol Biol 79:521

Zhao Y, Williams R, Prakash CS, He G (2012) Identification and characterization of gene-based SSR markers in date palm (*Phoenix dactylifera* L.). BMC Plant Biol 12(1):237

Zheng Y, El-kott AF, Shaldoum F, Massoud D, Pan Y, ElBealy ER (2021) Alleviation of diabetes-induced hepatotoxicity by date palm hydroalcoholic extract in rat model; a biochemical, immunohistochemical and stereological study. Int J Morphol 1(2):2

Zihad SNK, Uddin SJ, Sifat N, Lovely F, Rouf R, Shilpi JA, Sheikh BY, Göransson U (2021) Antioxidant properties and phenolic profiling by UPLC-QTOF-MS of Ajwah, Safawy and Sukkari cultivars of date palm. Biochem Biophys Rep 25:100909

Zohary D, Hopf M (2000) Domestication of plants in the old world: the origin and spread of cultivated plants in West Asia, Europe, and the Nile Valley, 3rd edn. Oxford University Press, Oxford

Zouine J, El Hadrami I (2007) Effect of 2, 4-D, glutamine and BAP on embryogenic suspension culture of date palm (*Phoenix dactylifera* L.). Sci Hortic 112(2):221–226

Current Advances in Health-Related Compounds in Sweet Cherry (*Prunus avium* L.)

Alejandro Calle, Ana Wünsch, Jose Quero-García, and Manuel Joaquín Serradilla

Contents

A. Calle (✉)
Plant and Environmental Sciences, Clemson University, Clemson, SC, USA
e-mail: acallec@clemson.edu

A. Wünsch
Centro de Investigación y Tecnología Agroalimentaria de Aragón, Zaragoza, Spain

Instituto Agroalimentario de Aragón-IA2 (CITA-Universidad de Zaragoza), Zaragoza, Spain
e-mail: awunsch@aragon.es

J. Quero-García
UMR Biologie du Fruit et Pathologie, INRAE, Univ. Bordeaux, Villenave d'Ornon, France
e-mail: jose.quero-garcia@inrae.fr

M. J. Serradilla
Centro de Investigaciones Científicas y Tecnológicas de Extremadura (CICYTEX), Instituto Tecnológico Agroalimentario de Extremadura (INTAEX), Badajoz, Spain
e-mail: manuel.serradilla@juntaex.es

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_38

Abstract

Sweet cherry is a highly appreciated fruit by the consumer because of its attractiveness and flavor. But it is also a nutritious fruit and an excellent source of bioactive compounds associated with health benefits. These molecules include nitrogenous and phenolic compounds like melatonin, serotonin, anthocyanins, or hydroxycinnamic acids. Numerous studies have associated their consumption with health-related properties, such as blood pressure regulation, anti-inflammatory properties, or cancer prevention. The concentration of these compounds has been evaluated in several cultivars, and a wide range of variability has been reported, which has been used as a starting point for genetic and molecular studies that have allowed identifying QTLs and candidate genes associated with bioactive compound concentration. At the same time, more recent transcriptomic and metabolomic studies are providing light into understanding their regulation and role during fruit development. Agricultural and postharvest practices are also showing useful to improve bioactive content. In this chapter, most relevant results of these works are reviewed.

Keywords

Sweet cherry · *Prunus avium* · Bioactive compounds · Biofortification · Health

1 Introduction: Sweet Cherries and Health

Sweet cherry (*Prunus avium* L.) is a temperate climate fruit tree species with a high commercial value in national and international markets due to its attractive appearance, delicious taste, and rich nutritional value, which is related to the presence of vitamins, fiber, minerals, fatty acids, and sugars (Gonçalves et al. 2017; Serradilla et al. 2017). Additionally, sweet cherries contain other secondary metabolites with high biological value, like flavonoids such as anthocyanins, hydroxycinnamic and hydroxybenzoic acids, flavanones, flavonols, and flavan-3-ols (Ballistreri et al. 2013; Serradilla et al. 2016; Gonçalves et al. 2017; Martini et al. 2017). These compounds exhibit high antioxidant activity that helps in combating cell damage, reducing inflammation, and promoting overall health (McCune et al. 2011; Martini et al. 2017). Therefore, cherry consumption is linked to beneficial and health-promoting effects (Nawirska-Olszańska et al. 2017). Specifically, sweet cherry consumption exhibits a reduced risk of gout and arthritis attacks and a reduction in gout-related pain (Singh et al. 2015). On the other hand, it also exhibits other health-related properties linked to blood pressure reduction, body weight control, diabetes, and the prevention of degenerative diseases like Alzheimer's (Wu et al. 2014; Kent et al. 2016; Gonçalves et al. 2017). Additionally, the health effects of sweet cherry

consumption and its derivatives have focused on physiological aspects related to anti-inflammatory and anticancer activities, protection against cardiovascular diseases, anti-obesity and antidiabetic activities, and neuroprotective factors (Blando and Oomah 2019). Moreover, these beneficial effects of sweet cherries are associated with the anti-inflammatory capacity shown by anthocyanins, inhibiting the production of nitric oxide and other pro-inflammatory factors, as well as improving vision, brain function, and sleep quality (Tsuda 2012; Garrido et al. 2013). Additionally, the consumption of sweet and sour cherry has been associated with regulating the sleep-wake cycle, mood performance, sleepiness, and jet lag symptoms, thanks to molecules like melatonin (González-Gómez et al. 2009; Howatson et al. 2012). The use of biostimulants or elicitors, both preharvest and postharvest, with salicylic acid and its derivatives, methyl jasmonate, oxalic acid, and, recently, melatonin, has been found to increase the content of these bioactive compounds that can improve fruit quality and also bioactive capacity (Yao and Tian 2005; Giménez et al. 2014, 2015; Martínez-Esplá et al. 2014; Sharafi et al. 2021; Carrión-Antolí et al. 2022).

2 Chemical Composition

The scientific community has extensively studied the chemical composition of cherries (Usenik et al. 2008; Correia et al. 2017; Serradilla et al. 2017; Papapetros et al. 2018). Cherries are characterized by a low-calorie level (around 63.0 kcal [263.34 kJ] per 100 g), high water content (approximately 80%), and the absence of sodium (Serradilla et al. 2017; Gonçalves et al. 2017). Water content varies according to genotype and ripening stage, ranging from 75% offered by the late ripening cultivar 'Starkrimson' to 88% shown by the mid-season ripening cultivar 'Sunburst' (Gonçalves et al. 2021). Regarding inorganic matter or ash content, values from 0.44%, shown by the mid-season ripening cultivar 'Skeena', to 2.88%, measured by the early ripening cultivar 'Burlat', were reported (Gonçalves et al. 2021).

2.1 Sugar Content

One of the parameters that have the most significant influence on consumer acceptance is sweetness (Crisosto et al. 2003; Serradilla et al. 2016). This trait depends on the soluble solid content (SSC) and soluble sugars. This content is influenced not only by genotype but also by agronomic practices, climatic conditions of the production area, and even rootstock (Hrotko 2008; Usenik et al. 2010; Correia et al. 2017; Serradilla et al. 2017). A broad range of SSC has been reported in the literature depending on cultivar and geographic localization. In the northern hemisphere, this range varies from 11.9 °Brix in the cultivar 'Celeste' to 24.5, led by the cultivar 'Salmo' (Girard and Kopp 1998; Gonçalves et al. 2021). Similarly, in the southern hemisphere, values ranging from 16.8 ('Santina') to 23.9 °Brix ('Bing') have been described (Param and Zoffoli 2016). On the other hand, SSC tends to

increase during fruit ripening (Serradilla et al. 2012). Cherries are characterized by different soluble sugars, such as disaccharides like sucrose, maltose, and trehalose and monosaccharides or reducing sugars like fructose, glucose, and especially D-glucose (Serradilla et al. 2017; Chen et al. 2022). In addition, other sugars, or polyols, such as sorbitol, are also present in cherries (Blando and Oomah 2019). The glucose content ranges from 6 to 10 g/100 g of fresh weight (FW), while the fructose content ranges from 4.6 to 6.7 g/100 g of FW. These values are influenced by genotype, rootstock, soil, climatic conditions, and fruit ripening stage (Serradilla et al. 2017).

Like other stone fruits, the ripening process in cherries is characterized by a double sigmoid curve (Serrano et al. 2005), where three stages of development are identified. The first stage is mainly based on cell division and elongation when the fruit is entirely immature or green; the second stage is the endocarp hardening, which leads to the formation of the nucleus and fruit color appearance; the third stage is the period of exponential growth caused by cell enlargement, in which physiological and biochemical changes in sugar, organic acid, and color occur drastically (Blando and Oomah 2019; Chen et al. 2022). During this process, soluble sugars show a trend consistent with changes exhibited by SSC: an initial increase, followed by stabilization, and a further increase in the last stage of the ripening process (Chen et al. 2022). Generally, during ripening, disaccharides are cleaved into their monomers, such as glucose and fructose. The latter two tend to accumulate during fruit ripening, while sorbitol concentration does not show any significant changes at the end of ripening (Teribia et al. 2016; Serradilla et al. 2017; Chen et al. 2022). All these changes during fruit development result from the joint action of multiple metabolites and genes (Yang et al. 2021a).

2.2 Organic Acids and Total Acidity

The sweet cherry flavor is mainly defined by sweetness and sourness (Serradilla et al. 2017). Sourness is expressed as titratable acidity (TA), which depends on genotypes and ripening stages. In the sweet cherry germplasm, TA ranges from 0.48 (‘Garnet’) to 0.96%, shown by ‘4–84’ and ‘Sweetheart’ cultivars (Gonçalves et al. 2021), reaching maximum values of TA at the end of the ripening process (Teribia et al. 2016). Cherries are generally considered medium acidity fruits, with pH ranging from 3.49 to 4.5 depending on cultivars and ripening stage (Ballistreri et al. 2013; Teribia et al. 2016; Gonçalves et al. 2021). Organic acids contribute to the flavor and pH of the fruits and, therefore, their sensory characteristics (Serradilla et al. 2017; Yang et al. 2021a). Among the primary organic acids found in cherries, malic acid, a predominant acid in the *Prunus* genus, stands out. Malic acid concentration depends on the cultivar and varies between 300 and 1100 mg/100 g of FW (Usenik et al. 2008; Serradilla et al. 2016); thus, some cultivars are more acidic than others. The following most crucial organic acids are citric acid, with values between 5 and 300 mg/100 g of FW, and succinic acid, which varies from 4 to 32 mg/100 g of FW (Serradilla et al. 2016). Other organic acids like oxalic, fumaric, and shikimic acid

have also been identified, showing concentrations below 8 mg/100 g of FW (Usenik et al. 2008; Serradilla et al. 2016). Like soluble sugars, malic acid content decreases at the beginning of ripening and then increases at the end of the process (Serrano et al. 2005). The balance of organic acids and sugars, defined by the SSC/TA ratio, affects cherry flavor and, consequently, customer acceptance (Crisosto et al. 2003; Serradilla et al. 2017). Gonçalves et al. (2021) found that this ratio ranged from 17.07 ('Sweetheart') to 32.33 ('Starkrimson'). However, Picota-type cultivars that show peduncle excision and late ripening time, like 'Ambrunés', can reach values close to 40 (Serradilla et al. 2012).

2.3 Nitrogenous Compounds

Cherries have also acquired great importance because their composition contains nitrogenous compounds of great bioactive value, such as indolamines, including melatonin (N-acetyl-5-methoxytryptamine), serotonin (5-hydroxytryptamine), and its precursor, the amino acid tryptophan. Within them, melatonin, a signaling molecule that regulates a wide range of physiological processes in plants and human circadian cycles, is a relevant compound in cherries (González-Gómez et al. 2009; Carrión-Antolí et al. 2022). Both melatonin and serotonin are synthesized from the essential amino acid tryptophan via the shikimic acid pathway. Similarly to other compounds, the melatonin and serotonin content depends on the cultivar and ripening stage of the fruits. For melatonin, concentration ranged from 0.6, shown by the Picota-type cultivar 'Pico Negro Limón', to 22.4 ng/100 g of FW of the early ripening cultivar 'Burlat'. Likewise, serotonin concentration ranges from 2.8 ('Pico Negro') to 37.6 ng/100 g of FW in 'Ambrunés'. Regarding the impact of the ripening stage, melatonin increased during ripening, while serotonin showed a less defined behavior (González-Gómez et al. 2009).

2.4 Phenolic Compounds

In addition to nitrogenous compounds, cherries, like other fruits, have acquired great importance in recent years due to the great content of bioactive compounds of phenolic nature that present high antioxidant activity against free radicals, which are related to beneficial properties for human health (Serra et al. 2011; Chockchaisawasdee et al. 2016; Correia et al. 2017; Serradilla et al. 2017; Blando and Oomah 2019; Gonçalves et al. 2021). The content of total phenols shown by cherries varies from 44.3 to 192 mg/100 g of FW depending on cultivar, ripening stage, and soil and climatic conditions (Serradilla et al. 2016). Among phenolic compounds present in cherries, flavonoids (anthocyanins, flavonols, and flavan-3-ols) and phenolic acids (hydroxycinnamic acids and hydroxybenzoic acids) are the most abundant (Fig. 1) (Serra et al. 2011; Ballistreri et al. 2013; Serradilla et al. 2017). Within phenolic acids, cherries stand out for hydroxycinnamic acids, especially neochlorogenic and p-coumaroylquinic acids, although chlorogenic and

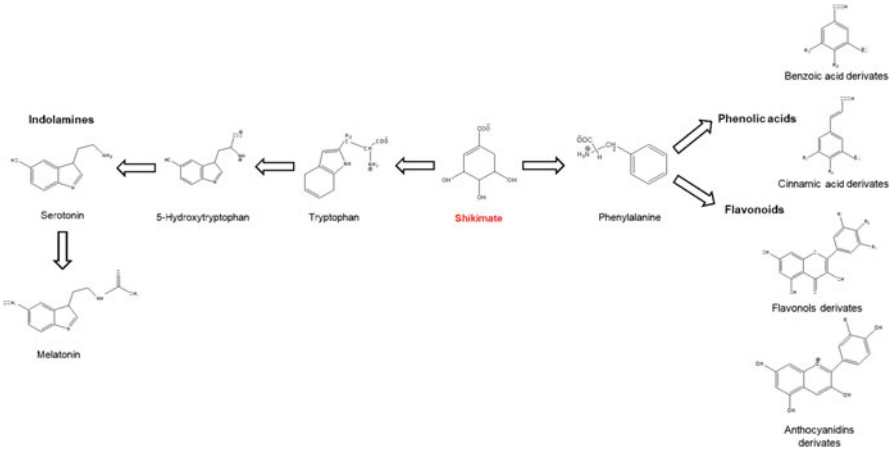


Fig. 1 Main phenolic compounds and indolamines present in sweet cherry fruits derived from shikimic acid (red) via different metabolic pathways

caffeoylquinic acids were detected in lower amounts (Serradilla et al. 2017). The concentration of these compounds is defined by genotype. For neochlorogenic acid, cultivars are classified based on concentration as low (4–20 mg/100 g of FW) like 'Burlat', 'Lapins', and 'Sweetheart'; medium (20–40 mg/100 g of FW) like 'Ferrovia', 'Blaze Star', and '0900 Ziraat'; and high (40–128 mg/100 g of FW; 'Bing') (Ballistreri et al. 2013). For p-coumaroylquinic acid content, intervals from 0.77 ('Lapins') to 131.45 ('Sam') mg/100 g of FW were reported in the bibliography (Serradilla et al. 2016). Moreover, the ratio of these two primary hydroxycinnamic acids is characteristic of each genotype (Mozetič et al. 2006).

Among flavonoids, cherries stand out for their high content of anthocyanins. These pigments are responsible for skin and flesh color and, therefore, for one of the most critical organoleptic characteristics that define consumer purchase (Serradilla et al. 2017; Gonçalves et al. 2021; Carrión-Antolí et al. 2022). Anthocyanins are characterized as water-soluble pigments present in the vacuole. They have a basic structure called aglycone or anthocyanidin (flavylium 2-phenyl-benzopyrylium) to which one or more sugars are attached by glycosidic bonds forming anthocyanins (Serradilla et al. 2016). The main anthocyanin identified in cherry has been cyanidin 3-*O*-rutinoside followed by cyanidin 3-*O*-ruthinoside and, in lower concentrations, peonidin 3-*O*-rutinoside, peonidin 3-*O*-glucoside, pelargonidin 3-*O*-rutinoside, and delphinidin 3-*O*-rutinoside (Serra et al. 2011; Ballistreri et al. 2013; Serradilla et al. 2017; Gonçalves et al. 2021). The cyanidin 3-*O*-rutinoside represents between 77% and 96% of the total anthocyanins, with light-colored cultivars showing average anthocyanin values between 2 and 47 mg/100 g of FW, while dark cherries can reach up to 297 mg/100 g of FW (Serradilla et al. 2016). Anthocyanins tend to accumulate during fruit ripening (Serrano et al. 2005; Serradilla et al. 2011).

Other flavonoids present in cherries, considered as noncolored phenolic compounds, are flavonols, remarkably quercetin 3-*O*-rutinoside or rutin, which concentration ranges

from 1.84 ('Van') to 51.97 mg/100 g of FW led by cultivar 'Lapins' (González-Gómez et al. 2010; Martini et al. 2017). Rutin content also increases during fruit ripening (Serradilla et al. 2011). Recently, Gonçalves et al. (2021) identified other flavonols such as quercetin 3-*O*-hexoside, kaempferol 3-*O*-rutinoside, kaempferol hexoside (3-gluc), and the kaempferol hexoside derivative 2, with concentrations strongly influenced by genotype. Other flavonoids listed as noncolored phenolic compounds are flavan-3-ols. In cherries, mainly epicatechin and catechin have been identified but also polymeric procyanidin (Blando and Oomah 2019). These represent less than 11.29% of the total noncolored compounds (Gonçalves et al. 2021), and their ratio is strongly influenced by genotype (Serra et al. 2011). Cherries are notorious for their high epicatechin content compared to catechin content, although this ratio also varies depending on the cultivar (Serra et al. 2011). Epicatechin levels range from 0.43 to 13.38 mg/100 g of FW, while catechin content ranges from 2.92 to 9.03 mg/100 g of FW (Serradilla et al. 2016). However, in the case of polymeric procyanidins, found in a few cultivars, levels of 20 mg/100 g of FW have been described, as is the case of the cultivar 'Royal Ann' (Chaovanalikit and Wrolstad 2004).

3 Genetics and QTL Analyses

In the last decade, several quantitative trait loci (QTLs) determining the genetic variation of the most relevant agronomic and fruit quality traits have been identified in sweet and sour cherry (Sooriyapathirana et al. 2010; Zhang et al. 2010; Dirlewanger et al. 2012; Rosyara et al. 2013; Castède et al. 2014; Quero-García et al. 2014; Campoy et al. 2015; Cai et al. 2019; Quero-García et al. 2019, 2021; Calle et al. 2020, 2021; Calle and Wünsch 2020; Branchereau et al. 2022; reviewed in Quero-García et al. 2022). Most of these works focused on physical traits like size, weight, firmness, and color (Sooriyapathirana et al. 2010; Zhang et al. 2010; Rosyara et al. 2013; Quero-García et al. 2014; Calle et al. 2020; Calle and Wünsch 2020). Despite the variation of nutritional and health-related compounds is well documented in sweet cherries during harvest and postharvest, as described above, few studies have investigated the genetics or carried QTL analyses of these compounds in sweet cherries (Quero-García et al. 2019; Calle et al. 2021).

3.1 Sugar Content and Total Acidity

Three works have focused on the genetics of sugar content and acidity in sweet cherry (Zhao et al. 2014; Quero-García et al. 2019; Calle and Wünsch 2020). These studies investigated sugar content, evaluated as SSC using a refractometer, TA, and pH (the latter only in Quero-García et al. 2019). Zhao et al. (2014) used 3-year phenotypic data from 601 pedigree-related individuals that included F_1 seedlings, breeding parental and commercial cultivars in a combined way for QTL analyses. Two major QTLs were identified on linkage groups (LGs) 2 and 4, with a minor QTL on LG7 for SSC. Quero-García et al. (2019) used an F_1 population

('Regina' × 'Garnet', N = 117) and 3-year data. This work identified SSC QTLs on LGs 1 and 3 by multi-year analysis, explaining from 6 to 10% of the variation. Later, in the multifamily QTL analysis conducted by Calle and Wünsch (2020), including 411 individuals from six populations of related pedigree and 2-year data, a major QTL for SSC, explaining nearly 30% of the phenotypic variation, was mapped on LG4. This QTL region overlapped with maturity and firmness QTLs on the same plant material. Also, a positive correlation among these traits was observed, with the latter maturing genotypes presenting a higher SSC value (Calle and Wünsch 2020). The results suggested the near presence of genes regulating sugar content and maturity, or genes with pleiotropic effects on these traits in this genome region (Calle and Wünsch 2020). In the same work, another minor SSC QTL was mapped on LG3, overlapping with the QTLs in the same region detected previously by Quero-García et al. (2019). For fruit acidity, high variability in QTLs detected between environmental years and populations was reported in these works, with QTLs on LGs 1, 2, 4, and 6 (Zhao et al. 2014; Quero-García et al. 2019; Calle and Wünsch 2020). Of them, only the genomic region at chromosome 6 seems stable across environments and plant materials for TA regulation in sweet cherry.

3.2 Fruit Color

Initial studies of the genetics of health-related compounds in sweet cherry were carried out indirectly by investigating fruit color. Sweet cherry color varies from yellow to mahogany or almost black, depending on the anthocyanin content profile (Gao and Mazza 1995; Jin et al. 2016; Calle et al. 2021). Fruit color was initially investigated as a relevant trait for sweet cherry breeding since different color fruits are preferred in different consumer markets (Crisosto et al. 2003). Furthermore, fruit color segregates in many breeding populations (Schmidt 1998). The cross of two red cultivars may result in blushed to dark cherries, and marker-assisted selection of this trait is desirable for breeders. By analyzing color segregation in sweet cherry progenies, the earliest studies showed that fruit skin color is determined by a major gene, in which alleles for red color are dominant to yellow or blushed (Fogle 1958; Schmidt 1998). The presence of minor genes exhibiting the epistasis effect was also proposed. More recently, research on the genetic control of skin and flesh color in sweet cherry by QTL analyses was carried out by Sooriyapathirana et al. (2010). In this work, fruit and skin color QTLs were mapped using a progeny from a cross of two cultivars with different flesh and skin color, blush-skinned and yellow-flesh 'Emperor Francis' and dark mahogany-skinned and red-fleshed 'New York 54' (Sooriyapathirana et al. 2010). The QTL analysis was conducted in 190 individuals that were phenotyped using a colorimeter, through a pseudo-testcross mapping strategy of a consensus map of 198 markers (102 SSR, 61 AFLP, 28 gene-derived markers, and seven SNP; Olmstead et al. 2008). A major QTL, explaining a large percentage of the phenotypic variation for skin and

flesh color, was identified on LG3. Also, minor QTLs were reported on LGs 6 and 8. This work confirmed that sweet cherry color was determined by a major regulatory gene (Sooriyapathirana et al. 2010). In a later study focused on deciphering the molecular mechanisms regulating fruit color, Jin et al. (2016) also mapped skin color using another sweet cherry F_1 segregating population. In this population ('Wanhongzhu' × 'Lapins'), both parents are dark red, and a 3:1 (dark red and blush fruits) segregation was observed for the 465 hybrids, confirming the regulation model by a major gene with dominance for dark red. In this case, specific-locus amplified fragment (SPAF) markers were used, and fruit color was mapped at a narrowed QTL (70.4–70.5 cM) on LG3. More recently, Calle et al. (2021) mapped skin and flesh color QTLs using 161 individuals of an F_1 population derived from the cross 'Vic' × 'Cristobalina' and genotyped with the RosBREED Cherry 6 + 9 K Illumina Infinium® SNP array (Vanderzande et al. 2020). The QTL analysis, using genetic maps of 910 ('Vic') and 789 ('Cristobalina') single nucleotide polymorphisms (SNPs) (Calle et al. 2021), also revealed a major QTL region for sweet cherry color on the same LG3 region previously reported by Sooriyapathirana et al. (2010) and Jin et al. (2016). This QTL region corresponds to chromosome 3, between 9.3–13.5 Mbp in *Prunus avium* v1.0.a1 (Shirasawa et al. 2017) and 11.1–15.4 Mbp of *Prunus avium* Tieton v1.0 genome (Wang et al. 2020). An additional minor QTL for flesh color was identified on LG4, explaining 15% of phenotypic variation when colorimeter data was used for QTL mapping (Calle et al. 2021).

3.3 Anthocyanins

QTL analyses of anthocyanin content in sweet cherry were also carried out by Calle et al. (2021) in the same work as previously reported for skin and flesh color QTLs. Anthocyanin compounds were extracted and quantified by high-performance liquid chromatography (HPLC) from 15 mature fruits of each progeny hybrid. Four anthocyanins (cyanidin 3-*O*-glucoside, cyanidin 3-*O*-rutinoside, peonidin 3-*O*-glucoside, and peonidin 3-*O*-rutinoside) were identified and quantified. The use of these data for QTL mapping revealed the most significant QTLs for the four anthocyanins on the same genomic region of chromosome 3 (Fig. 2) where fruit color had been mapped by Sooriyapathirana et al. (2010), Jin et al. (2016), and in the same population (Calle et al. 2021). The percentage of phenotypic variation explained by these QTLs ranged from 12 to 23%, and these results confirmed fruit color and anthocyanin biosynthesis have the same genetic determinants. In addition to this main QTL region on LG3, other minor anthocyanin content QTLs were also mapped on LG7 for cyanidin 3-*O*-glucoside, cyanidin 3-*O*-rutinoside, and peonidin 3-*O*-glucosideexplaining up to 10% of phenotypic variation and on LG4 for peonidin 3-*O*-rutinoside (Fig. 2) in the same region where a color QTL was also mapped in the same population (Calle et al. 2021). These results indicate that beyond a major gene in chromosome 3, other genes in chromosomes 4 and 7 also contribute, to a lesser extent, to anthocyanin content regulation in sweet cherry.

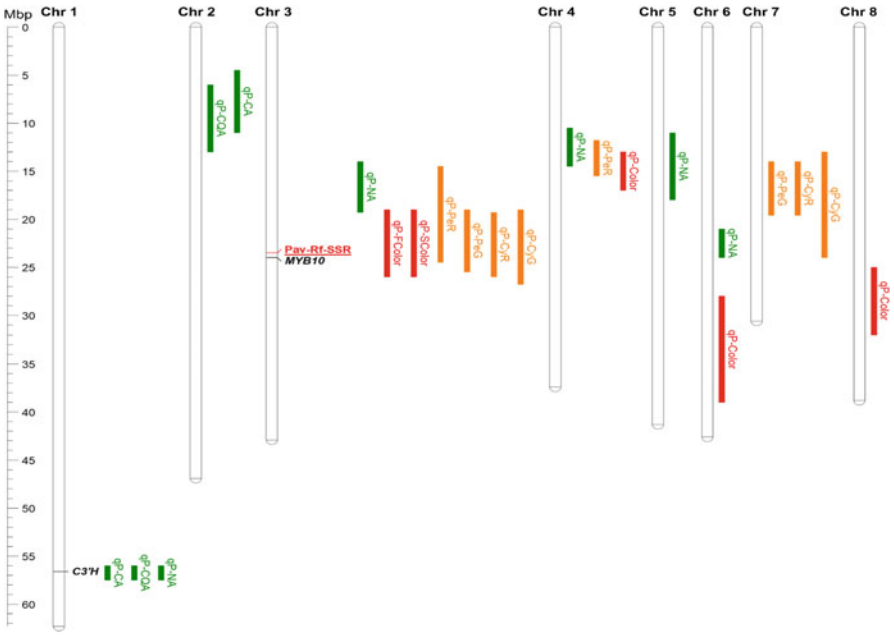


Fig. 2 Physical position in the *Prunus avium* L. genome ('Tieton' v2.0; Wang et al. 2020) of QTLs (Sooriyapathirana et al. 2010; Jin et al. 2016; Calle et al. 2021) associated with health-related (HR) compounds identified in the species. Fruit color QTLs (*qP-Color*) are shown in red, anthocyanin content QTLs (*qP-CyG*, cyanidin 3-O-glucoside; *qP-CyR*, cyanidin 3-O-rutinoside; *qP-PeG*, peonidin 3-O-glucoside; and *qP-PeR*, peonidin 3-O-rutinoside) in orange, and hydroxycinnamic acid QTLs (*qP-NA*, neochlorogenic acid; *qP-CQA*, *ρ*-coumaroylquinic acid; *qP-CA*, *ρ*-coumaric acid) in green. Position of candidate genes associated with HR-compound biosynthesis (*C3'H*, coumarate 3-hydroxylase and *MYB10*) are also included (Sooriyapathirana et al. 2010; Calle et al. 2021). Physical position of Pav-Rf-SSR marker for fruit color prediction is also included (Sanderful et al. 2016)

3.4 Hydroxycinnamic Acids

The genetics of hydroxycinnamic acid content was also investigated in sweet cherry by Calle et al. (2021). Three compounds, neochlorogenic acid, p-coumaroylquinic acid, and p-coumaric acid, were also detected and quantified by HPLC and analyzed for QTL mapping using the same population and maps as for color and anthocyanin content in Calle et al. (2021). In this case, a major QTL was detected at the bottom region of LG1 (Fig. 2) for the three compounds (141.34–141.63 cM; 46.67–47.17 Mbp in the peach genome v2.0.a1). These QTLs were highly significant (logarithm of the odds: LOD $>$ 20) and explained a substantial percentage of the phenotypic variation (60–78%; Calle et al. 2021). Other less significant QTLs were found on LGs 3, 4, 5, and 6. These were associated with minor variations in the concentration of these compounds (4–11% of phenotypic variation). Similarly, two additional minor QTLs for p-coumaroylquinic acid and p-coumaric acid were found on the same region of LG2

(Fig. 2) (2.24–6.09 Mbp; peach genome v2.0.a1) (Calle et al. 2021). The results of this work showed that a major gene also regulates the concentration of these compounds and is likely located at the bottom of chromosome 1.

4 Candidate Genes

4.1 Sugars and Organic Acids

Genetic regulation of sugars and organic acids in cherries remains largely uncharacterized. Only a few studies investigated changes during fruit development related to the biosynthesis of these compounds (Beaver et al. 1995; Gao et al. 2003; Yang et al. 2021a). Initial studies in sour cherry revealed two *sorbitol transporter* genes (*SOT1* and *SOT2*) that were expressed during fruit development playing a major role in sorbitol accumulation in the species (Gao et al. 2003). In a more recent study by Yang et al. (2021a), transcriptomic analysis during the fruit development period of the sweet cherry cultivar 'Black Pearl' revealed genes like *SOT*, *PFK* (phosphofructokinase), *NINV* (neutral invertase), or *SUS* (sucrose synthase) highly upregulated during early stages of fruit development and another set of genes like *PGI* (phosphoglucoisomerase), *UGP* (UDPG-pyrophosphorylase), *PGM* (phosphoglucomutase), and *SPS* (sucrose-phosphate synthase) upregulated during the late stage of fruit development when sucrose accumulation was enhanced. Six additional genes of the *SWEET* family were identified, showing different expression patterns related to sugar transports (Yang et al. 2021a). Regarding organic acids, just a few genes like *PEPC* (phosphoenolpyruvate carboxylase) and one *MDH* (malate dehydrogenase) that were downregulated during fruit development were observed to keep acid content low by decreasing malate acid during cherry development (Yang et al. 2021a). In this same study, other genes responsible for the metabolism of organic acids like *GOGAT* (glutamate synthase) and *GDH* (glutamate dehydrogenase) increased their expression during fruit development, reaching the highest transcription close to harvest time (Yang et al. 2021a).

4.2 Fruit Color and Anthocyanins

Most of the genetic and molecular studies involving health-related compounds in sweet and sour cherries focused on anthocyanin content regulation. Transcription factors (TFs) are known to be critical regulators of anthocyanin biosynthesis in cherries (Lin-Wang et al. 2010; Jin et al. 2016; Chen et al. 2022). The complexes of TFs containing an R2R3 MYB, a WD40, and a basic helix-loop-helix (bHLH) were observed in other species to promote anthocyanin accumulation through binding to genes involved in the flavonoid pathway (Allan et al. 2008). In sweet cherry, various studies indicated that the R2R3-MYB TF *PavMYBA/PavMYB10* interacts with two other TFs (*PavbHLH3* and *PavWD40*) to form an MBW complex (MYB-bHLH-WD40) that regulates the expression of two key genes associated with the

anthocyanin pathway (*PavANS* and *PavUFGT*) (Shen et al. 2014; Starkevič et al. 2015; Jin et al. 2016; Guo et al. 2018). This MYB TF (*MYB10*) is located at the primary color, and anthocyanin QTLs reported in sweet cherry chromosome 3 (Sooriyapathirana et al. 2010; Jin et al. 2016; Calle et al. 2021), supporting its relevance. This transcription factor was first cloned by Lin-Wang et al. (2010) in sweet and sour cherry. Based on additional expression analysis by qPCR, a clear correlation between anthocyanin content and gene expression was confirmed (Lin-Wang et al. 2010). Later works found the same correlation in the dark red cultivar 'Hong Deng' (Shen et al. 2014). These studies confirmed *PavMYB10* as one of the key genes associated with variation in flavonoid content in the species, and it was the basis for developing a DNA test (Pav-Rf-SSR) to predict fruit color in sweet cherry (Fig. 2) (Sanderful et al. 2016). Additionally, various alleles for *PavMYB10* were associated with different ranges of flavonoid accumulation and fruit coloration in sweet cherry (Sooriyapathirana et al. 2010; Jin et al. 2016). Other TFs belonging to 22 different families, specifically bZIP, C2C2-Dof, NAC, bHLH, and R2R3-MYB, were also observed to mediate in anthocyanin regulation in response to light in the 'Rainier' cultivar (Guo et al. 2018).

4.3 Hydroxycinnamic Acids

The study by Calle et al. (2021) suggested a candidate gene for the regulation of the main hydroxycinnamic acid content in sweet cherries. The gene, *ρ-coumarate 3-hydroxylase* (*C3H*), located within LG1 QTL confidence interval identified in the same work, was considered the strongest candidate gene for hydroxycinnamic acid regulation in sweet cherry (Calle et al. 2021). This enzyme was previously observed in other species and is involved in the hydroxycinnamic acid biosynthesis pathway (Chagné et al. 2012; Kim et al. 2019). The LG1 region in which sweet cherry had the most significant QTLs for hydroxycinnamic acids was found to be syntenic to LG15 in apple (Illa et al. 2011), where also a major QTL for hydroxycinnamic acid content had been identified in apple (Chagné et al. 2012), leading to the identification of this candidate gene by Calle et al. (2021) and suggesting a common mechanism for the regulation of these phenolic acids.

5 Gene Expression and Functional Analyses

In sweet and sour cherry, strong correlations between the expression of structural and regulatory genes have been associated with health-related compound biosynthesis pathways (Lin-Wang et al. 2010; Liu et al. 2013; Shen et al. 2014; Starkevič et al. 2015; Wei et al. 2015; Guo et al. 2018). Within health-related compounds, anthocyanin biosynthesis was the widest studied in cherries. These compounds are synthesized through the flavonoid pathway (Winkel-Shirley 2001), in which most of the genes related to the regulation and biosynthesis of anthocyanins are well characterized in model plants (Allan et al. 2008). Initial studies were conducted to find candidate genes associated with anthocyanin biosynthesis in cherries by real-time PCR (RT-PCR)

targeting some candidate genes reported in other species. Lin-Wang et al. (2010) analyzed the expression level of two genes (anthocyanidin synthase [*ANS*] and chalcone synthase [*CHS*]) potentially involved in the anthocyanin biosynthesis pathway through RT-PCR. Two cultivars, 'Stella' (red coloration) and 'Rainier' (bicolor coloration), were used for comparison of expression profiling during fruit development, in which expression of *ANS* and *CHS* genes showed upregulation correlated with anthocyanin content (Lin-Wang et al., 2010). Similarly, the expression level of six anthocyanin biosynthesis genes (*ANS*, *CHS*, chalcone isomerase [*CHI*], flavanone 3-hydroxylase [*F3H*], dihydroflavonol 4-reductase [*DFR*], and UDP glucose: flavonol 3-O-glucosyltransferase [*UFGT*]) was analyzed by RT-PCR in high-anthocyanin-content cultivars ('Hongdeng' and 'Caihong' by Liu et al. (2013); 'Kitayanka', 'Irema BS', 'Werdersche braune', and 'Belobokaya rannyaya' by Starkevič et al. (2015)). In these studies, as previously reported by Lin-Wang et al. (2010), all investigated gene expression manifested a significant correlation with anthocyanin content, in which the expression level of these genes increases through fruit development.

Whole genome transcriptome analyses have complemented these findings. Wei et al. (2015) sequenced and annotated a reference transcriptome of sweet cherry to identify genes associated with anthocyanin biosynthesis. In this study, they sampled total anthocyanin content during four stages of fruit development (from 20 to 55 days after pollination) and sequenced the transcriptome of yellow ('13–33') and dark red ('Tieton') cultivars for comparison. The biological pathway analysis of these transcripts revealed 72 genes related to anthocyanin biosynthesis during fruit development, including key genes like *ANS*, *CHS*, *CHI*, *F3H*, *DFR*, *UFGT*, phenylalanine ammonia-lyase (*PAL*), 4-coumarate-CoA ligase (*4CL*), and flavanone 3′-hydroxylase (*F3'H*) involved in anthocyanin pathway (Wei et al. 2015). As observed in the RT-PCR studies, these genes were highly upregulated in the dark red cultivar ('Tieton') and downregulated in the yellow fruits of '13–33' (Wei et al. 2015). Guo et al. (2018) also performed RNA sequencing (RNAseq) analysis of anthocyanin biosynthesis in sweet cherry to investigate the anthocyanin accumulation mechanism further. The transcriptomes of two contrasting cultivars, 'Hongdeng' (dark) and 'Rainier' (bicolor), were sequenced during fruit development, and as observed in previous studies in cherries (Starkevič et al. 2015; Wei et al. 2015; Shen et al. 2014; Lin-Wang et al. 2010), genes encoding enzymes like *CHS*, *CHI*, *F3H*, and *F3'H* were significantly upregulated in dark fruits. Similarly, results were reported by Yang et al. (2021b) in which weighted gene co-expression network analysis (WGCNA) indicated similar expression patterns of *4CL2* and *ANS* with 11 TF (*bHLH13/74*, *DIV*, *ERF109/115*, *GATA8*, *GT2*, *GTE10*, *MYB308*, *PosF21*, and *WRKY7*). Additionally, another set of genes related to hormone and light signaling, like *protein phosphatase 2Cs*, *PHYTOCHROME INTERACTING FACTOR 3* (*PIF3*), *phytochromes*, and *ELONGATED HYPOCOTYL 5* (*HY5*), were also significantly differentially expressed between dark- and light-colored fruits and therefore associated with anthocyanin biosynthesis (Guo et al. 2018). Together, gene expression analyses have shown significant differences in transcript accumulation between cultivars with high and low anthocyanin accumulation, which are especially significant in expression for *UFGT* and *CHS* (Shen et al. 2014; Starkevič et al. 2015; Wei et al. 2015).

In addition to the genes directly involved in the anthocyanin pathway, various studies suggested that genes related to the signaling pathways of phytohormones like abscisic acid (ABA) and gibberellic acid (GA) play essential roles in the accumulation of health-related compounds (Sun et al. 2010; Jia et al. 2011; Guo et al. 2018; Chen et al. 2022). Shen et al. (2014) reported that silencing a gene (*PavNCED1*) essential for ABA biosynthesis (Zhang et al. 2009) blocked anthocyanin biosynthesis in sweet cherry, suggesting a complex interaction of genes regulating anthocyanin pathway beyond enzymes involved in the flavonoid pathway. In the same way, Guo et al. (2018) identified 32 genes required for plant hormone biosynthesis that were regulated in response to light exposure. In this study, differentially expressed genes related to ABA, GA, auxin, and ethylene pathways were modulated by light, being ABA and GA the only phytohormones that correlated with expression profiles of *HY5* and *PIF3* genes, suggesting that ABA and GA might be the main light-dependent hormones associated with anthocyanin biosynthesis (Guo et al. 2018).

6 Future Prospects and Conclusions

The works reviewed here show that sweet cherries are a health-valuable fruit, which is also highly appreciated by consumers, including children, making them an excellent source of health-related compounds. Therefore, it is important that the healthy quality of sweet cherries is known and communicated so it can be further consumed and produced. At the same time, sweet cherries are highly perishable and seasonal, which complicates their availability throughout the year and to markets far from the cultivation area. Agronomic practices, breeding, and postharvest technology research are allowing to improve these drawbacks, and sweet cherry production, quality, and availability (in time and space) have largely improved in the last years. Extensive efforts are being made in these areas, which have resulted in remarkable improvement, but it is essential not to lose the healthy qualities of sweet cherry in the way. In fact, breeding for a healthier cherry is still lacking, and a more extensive evaluation of HR compounds in sweet cherry collections would be needed. As healthy nutrition is more evident every day and society more aware of it, it is necessary to move forward to obtain a healthier fruit, if possible, or a designed fruit that can help mitigate diet deficiencies or fight specific diseases. The availability of new technologies will help to move forward in this direction. In this sense, agronomic and postharvest practices are useful in improving sweet cherries' health qualities.

References

Allan AC, Hellens RP, Laing WA (2008) MYB transcription factors that colour our fruits. Trends Plant Sci 13:99–102. https://doi.org/10.1016/j.tplants.2007.11.012

Ballistreri G, Continella A, Gentile A, Amenta M, Fabroni S, Rapisarda P (2013) Fruit quality and bioactive compounds relevant to human health of sweet cherry *(Prunus avium* L.) cultivars grown in Italy. Food Chem 140:630–638. https://doi.org/10.1016/j.foodchem.2012.11.024

Beaver JA, Iezzoni AF, Ramm CW (1995) Isozyme diversity in sour, sweet, and ground cherry. Theor Appl Genet 90:847–852. https://doi.org/10.1007/BF00222021

Blando F, Oomah BD (2019) Sweet and sour cherries: origin, distribution, nutritional composition and health benefits. Trends Food Sci Technol 86:517–529. https://doi.org/10.1016/j.tifs.2019.02.052

Branchereau C, Quero-García J, Zaracho-Echagüe NH, Lambelin L, Fouché M, Wenden B, Donkpegan A, Le Dantec L, Barreneche T, Alletru D, Parmentier J, Dirlewanger E (2022) New insights into flowering date in *Prunus*: fine mapping of a major QTL in sweet cherry. Hortic Res 9:uhac042. https://doi.org/10.1093/hr/uhac042

Cai L, Quero-García J, Barreneche T, Dirlewanger E, Saski C, Iezzoni A (2019) A fruit firmness QTL identified on linkage group 4 in sweet cherry (*Prunus avium* L.) is associated with domesticated and bred germplasm. Sci Rep 9:5008. https://doi.org/10.1038/s41598-019-41484-8

Calle A, Wünsch A (2020) Multiple-population QTL mapping of maturity and fruit-quality traits reveals LG4 region as a breeding target in sweet cherry (*Prunus avium* L.). Hortic Res 7:127. https://doi.org/10.1038/s41438-020-00349-2

Calle A, Balas F, Cai L, Iezzoni A, López-Corrales M, Serradilla MJ, Wünsch A (2020) Fruit size and firmness QTL alleles of breeding interest identified in a sweet cherry 'Ambrunés' × 'Sweetheart' population. Mol Breed 40:86. https://doi.org/10.1007/s11032-020-01165-1

Calle A, Serradilla MJ, Wünsch A (2021) QTL mapping of phenolic compounds and fruit colour in sweet cherry using a 6+9K SNP array genetic map. Sci Hortic 280:109900. https://doi.org/10.1016/j.scienta.2021.109900

Campoy JA, Le Dantec L, Barreneche T, Dirlewanger E, Quero-García J (2015) New insights into fruit firmness and weight control in sweet cherry. Plant Mol Biol Report 33(4):783–796. https://doi.org/10.1007/s11105-014-0773-6

Carrión-Antolí A, Martínez-Romero D, Guillén F, Zapata PJ, Serrano M, Valero D (2022) Melatonin pre-harvest treatments leads to maintenance of sweet cherry quality during storage by increasing antioxidant systems. Front Plant Sci 13:863467. https://doi.org/10.3389/fpls.2022.863467

Castède S, Campoy JA, Quero-García J, Le Dantec L, Lafargue M, Barreneche T, Wenden B, Dirlewanger E (2014) Genetic determinism of phenological traits highly affected by climate change in *Prunus avium*: flowering date dissection into chilling and heat requirements. New Phytol 202(2):703–715. https://doi.org/10.1111/nph.12658

Chagné D, Krieger C, Rassam M, Sullivan M, Fraser J, André C, Pindo M, Troggio M, Gardiner SE, Henry RA, Allan AC, McGhie TK, Laing WA (2012) QTL and candidate gene mapping for polyphenolic composition in apple fruit. BMC Plant Biol 12:12. https://doi.org/10.1186/1471-2229-12-12

Chaovanalikit A, Wrolstad RE (2004) Anthocyanin and polyphenolic composition of fresh and processed cherries. J Food Sci 69:107–117. https://doi.org/10.1111/j.1365-2621.2004.tb17859.x

Chen C, Chen H, Yang W, Li J, Tang W, Gong R (2022) Transcriptome and metabolomic analysis of quality changes during sweet cherry fruit development and mining of related genes. Int J Mol Sci 23(13):7402. https://doi.org/10.3390/ijms23137402

Chockchaisawasdee S, Golding JB, Vuong QV, Papoutsis K, Stathopoulos CE (2016) Sweet cherry: composition, postharvest preservation, processing and trends for its future use. Trends Food Sci Technol 55:72–83. https://doi.org/10.1016/j.tifs.2016.07.002

Correia S, Schouteh R, Silva AP, Gonçalves B (2017) Factors affecting quality and health promoting compounds during growth and postharvest life of sweet cherry (*Prunus avium* L.). Front Plant Sci 8:2166. https://doi.org/10.3389/fpls.2018.00464

Crisosto CH, Crisosto GM, Metheney P (2003) Consumer acceptance of 'Brooks' and 'Bing' cherries is mainly dependent on fruit SSC and visual skin color. Postharvest Biol Technol 28:159–167. https://doi.org/10.1016/S0925-5214(02)00173-4

Dirlewanger E, Quero-García J, Le Dantec L, Lambert P, Ruiz D, Dondini L, Illa E, Quilot-Turion B, Audergon JM, Tartarini S, Letourmy P, Arús P (2012) Comparison of the genetic determinism of two key phenological traits, flowering and maturity dates, in three *Prunus* species: peach, apricot and sweet cherry. Heredity 109:280–292. https://doi.org/10.1038/hdy.2012.38

Fogle HW (1958) Inheritance of fruit color in sweet cherries (*Prunus avium*). J Hered 49:294–298

Gao L, Mazza G (1995) Characterization, quantification and distribution of anthocyanins and colorless phenolic in sweet cherries. J Agric Food Chem 43:343–346. https://doi.org/10.1021/jf00050a015

Gao ZF, Maurousset L, Lemoine R, Yoo SD, Nocker SV, Loescher W (2003) Cloning, expression, and characterization of sorbitol transporters from developing sour cherry fruit and leaf skin tissue. Plant Physiol 131(4):1566–1575. https://doi.org/10.1104/pp.102.016725

Garrido M, González-Gómez D, Lozano M, Barriga C, Paredes SD, Rodríguez AB (2013) A Jerte valley cherry product provides beneficial effects on sleep quality. Influence on aging. J Nutr Health Aging 17:553–560. https://doi.org/10.1007/s12603-013-0029-4

Giménez MJ, Valverde JM, Valero D, Guillén F, Martínez-Romero D, Serrano M, Castillo S (2014) Quality and antioxidant properties on sweet cherries as affected by preharvest salicylic and acetylsalicylic acids treatments. Food Chem 160:226–232. https://doi.org/10.1016/j.foodchem.2014.03.107

Giménez MJ, Valverde JM, Valero D, Díaz-Mula HM, Zapata PJ, Serrano S (2015) Methyl salicylate treatments of sweet cherry trees improve fruit quality at harvest and during storage. Sci Hortic 197:665–673. https://doi.org/10.1016/j.scienta.2015.10.033

Girard B, Kopp TG (1998) Physicochemical characteristics of selected sweet cherry cultivars. J Agric Food Chem 46:471–476. https://doi.org/10.1021/jf970646j

Gonçalves AC, Bento C, Silva BM, Silva LR (2017) Sweet cherries from Fundão possess anti-diabetic potential and protect human erythrocytes against oxidative damage. Food Res Int 95:91–100. https://doi.org/10.1016/j.foodres.2017.02.023

Gonçalves AC, Campos G, Alves G, García-Viguera C, Moreno DA (2021) Silva LR (2021) physical and phytochemical composition of 23 Portuguese sweet cherries as conditioned by variety (or genotype). Food Chem 335:127637. https://doi.org/10.1016/j.foodchem.2020.127637

González-Gómez D, Lozano M, Fernández-León MF, Ayuso MC, Bernalte MJ, Rodríguez AB (2009) Detection and quantification of melatonin and serotonin in eight sweet cherry cultivars (*Prunus avium* L.). Eur Food Res Technol 229:223–229. https://doi.org/10.1007/s00217-009-1042-z

González-Gómez D, Lozano M, Fernández-León MF, Bernalte MJ, Ayuso MC, Rodríguez AB (2010) Sweet cherry phytochemicals: identification and characterization by HPLC-DAD/ESI-MS in six sweet-cherry cultivars grown in Valle del Jerte (Spain). J Food Compos Anal 23:533–539. https://doi.org/10.1016/j.jfca.2009.02.008

Guo X, Wang Y, Zhai Z, Huang T, Zhao D, Peng X, Feng C, Xiao Y, Li T (2018) Transcriptomic analysis of light-dependent anthocyanin accumulation in bicolored cherry fruits. Plant Physiol Biochem 130:663–677. https://doi.org/10.1016/j.plaphy.2018.08.016

Howatson G, Bell PG, Tallent J, Middleton B, McHugh MP, Ellis J (2012) Effect of tart cherry juice (*Prunus cerasus*) on melatonin levels and enhanced sleep quality. Eur J Nutr 51:909–916. https://doi.org/10.1007/s00394-011-0263-7

Hrotko K (2008) Progress in cherry rootstock research. Acta Hortic 795:171–178

Illa E, Sargent DJ, López-Girona E, Bushakra J, Cestaro A, Crowhurst R, Pindo M, Cabrera A, van der Knaap E, Iezzoni A, Gardiner S, Velasco R, Arús P, Chagné D, Troggio M (2011) Comparative analysis of rosaceous genomes and the reconstruction of a putative ancestral genome for the family. BMC Evol Biol 11:9. https://doi.org/10.1186/1471-2148-11-9

Jia HF, Chai YM, Li CL, Lu D, Luo JJ, Qin L (2011) Abscisic acid plays an important role in the regulation of strawberry fruit ripening. Plant Physiol 157:188–199. https://doi.org/10.1104/pp.111.177311

Jin W, Wang H, Li M, Wang J, Yang Ym Zhang X, Yan G, Zhang H, Liu J, Zhang K (2016) The R2R3 MYB transcription factor *PavMYB10.1* involves in anthocyanin biosynthesis and determines fruit color skin colour in sweet cherry (*Prunus avium* L.). Plant Biotechnol J 14(11):2120–2133. https://doi.org/10.1111/pbi.12568

Kent K, Charlton KE, Jenner A, Roodenrys S (2016) Acute reduction in blood pressure following consumption of anthocyanin-rich cherry juice may be dose-interval dependant: a pilot cross-over study. Int J Food Sci Nutr 67:47–52. https://doi.org/10.3109/09637486.2015.1121472

Kim YK, Swanson RT, Alvarez MI, Johnson TS, Cho KH, Clark DG, Colquhoun TA (2019) Down regulation of p-coumarate 3-hydroxylase in petunia uniquely alters the profile of emitted floral volatiles. Sci Rep 9:8852. https://doi.org/10.1038/s41598-019-45183-2

Lin-Wang K, Bolitho J, Grafton K, Kortstee A, Karunairetnam S, McGhie TK, Espley RV, Hellens RP, Allan AC (2010) An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. BMC Plant Biol 10:50. https://doi.org/10.1186/1471-2229-10-50

Liu Y, Shen X, Zhao K, Ben Y, Guo X, Zhang X, Li T (2013) Expression analysis of anthocyanin biosynthetic genes in different colored sweet cherries (*Prunus avium* L.) during fruit development. J Plant Growth Regul 32:901–907. https://doi.org/10.1007/s00344-013-9355-3

Martínez-Esplá A, Zapata PJ, Valero D, García-Viguera C, Castillo S, Serrano M (2014) Preharvest application of oxalic acid increased fruit size, bioactive compounds, and antioxidant capacity in sweet cherry cultivars (*Prunus avium* L.). J Agric Food Chem 62:3432–3437. https://doi.org/10.1021/jf500224g

Martini S, Conte A, Tagliazucchi D (2017) Phenolic compounds profile and antioxidant properties of six sweet cherry (*Prunus avium*) cultivars. Food Res Int 97:15–26. https://doi.org/10.1016/j.foodres.2017.03.030

McCune LM, Kubota C, Stendell-Hollins NR, Thomson CA (2011) Cherries and health: a review. Crit Rev Food Sci Nutr 51:1–12. https://doi.org/10.1080/10408390903001719

Mozetič B, Simčič M, Trebše P (2006) Anthocyanins and hydroxycinnamic acids of Lambert Compact cherries (*Prunus avium* L.) after cold storage and 1-methylcyclopropene treatment. Food Chem 97(2):302–309. https://doi.org/10.1016/j.foodchem.2005.04.018

Nawirska-Olszańska A, Kolniak-Ostek J, Oziembłowski M, Ticha A, Hyšpler R, Zadak Z, Židová P, Paprstein F (2017) Comparison of old cherry cultivars grown in Czech Republic by chemical composition and bioactive compounds. Food Chem 228:136–142. https://doi.org/10.1016/j.foodchem.2017.01.154

Olmstead JW, Sebolt AM, Cabrera A, Sooriyapathirana SS, Hammer S, Iriarte G, Wang D, Chen CY, van der Knaapp E, Iezzoni AF (2008) Construction of an intra-specific sweet cherry (*Prunus avium* L.) genetic linkage map and synteny analysis with the *Prunus* reference map. Tree Genet Genomes 4:897–910. https://doi.org/10.1007/s11295-008-0161-1

Papapetros S, Louppis A, Kosma I, Kontakos S, Badeka A, Kontominas MG (2018) Characterization and differentiation of botanical and geographical origin of selected popular sweet cherry cultivars grown in Greece. J Food Compos Anal 72:48–56. https://doi.org/10.1016/j.jfca.2018.06.006

Param N, Zoffoli JP (2016) Genotypic differences in sweet cherries are associated with the susceptibility to mechanical damage. Sci Hortic 211:410–419. https://doi.org/10.1016/j.scienta.2016.09.027

Quero-García J, Fodor A, Reignier A, Capdeville G, Joly J, Tauzin Y, Fouilhaux L, Dirlewanger E (2014) QTL detection of important agronomic traits for sweet cherry breeding. Acta Hortic 1020:57–64. https://doi.org/10.17660/ActaHortic.2014.1020.5

Quero-García J, Campoy JA, Barreneche T, Le Dantec L, Wenden B, Fouché M, Dirlewanger E, Silva H, Cai L, Iezzoni A (2019) Present and future of marker-assisted breeding in sweet and sour cherry. Acta Hortic 1235. https://doi.org/10.17660/ActaHortic.2019.1235.1

Quero-García J, Letourmy P, Campoy JA, Branchereau C, Malchev S, Barreneche T, Dirlewanger E (2021) Multi-year analyses on three populations reveal the first stable QTLs for tolerance to rain-induced fruit cracking in sweet cherry (*Prunus avium* L.). Hortic Res 8:136. https://doi.org/10.1038/s41438-021-00571-6

Quero-García J, Branchereau C, Barreneche T, Dirlewanger E (2022) DNA-informed breeding in sweet cherry: current advances and perspectives. Italus Hortus 29:14–37. https://doi.org/10.26353/j.itahort/2022.1.C2

Rosyara UR, Bink MCAM, van de Weg E, Zhang G, Wang D, Sebolt A, Dirlewanger E, Quero-García J, Shuster M, Iezzoni AF (2013) Fruit size QTL identification and the prediction of parental QTL genotypes and breeding values in multiple pedigreed populations in sweet cherry. Mol Breed 32:875–887. https://doi.org/10.1007/s11032-013-9916-y

Sanderful P, Oraguzie M, Peace C (2016) A DNA test for routine prediction in breeding of sweet cherry fruit color, Pav-Rf-SSR. Mol Breed 36:33. https://doi.org/10.1007/s11032-016-0458-y

Schmidt H (1998) On the basis of fruit color in sweet cherry. Acta Hortic 468:77–81

Serra AT, Duarte RO, Bronze MR, Duarte CMM (2011) Identification of bioactive response in traditional cherries from Portugal. Food Chem 125:318–325. https://doi.org/10.1016/j.foodchem.2010.07.088

Serradilla MJ, Lozano M, Bernalte MJ, Ayuso MC, López-Corrales M, González-Gómez D (2011) Physicochemical and bioactive properties evolution during ripening of 'Ambrunes' sweet cherry cultivar. LWT-Food Sci Technol 44:199–205. https://doi.org/10.1016/j.lwt.2010.05.036

Serradilla MJ, Martín A, Ruiz-Moyano S, Hernández A, López-Corrales M, Córdoba MG (2012) Physicochemical and sensorial characterisation of four sweet cherry cultivars grown in Jerte Valley (Spain). Food Chem 133:1551–1559. https://doi.org/10.1016/j.foodchem.2012.02.048

Serradilla MJ, Hernández A, López-Corrales M, Ruíz-Moyano S, Córdoba MG, Martín A (2016) Composition of the cherry (*Prunus avium* L. and k L.; Rosaceae). In: Simmonds MSJ, Preedy VR (eds) Nutritional compositions of fruit cultivars (pp 127–147). Academic Press, Cambridge, Massachusetts, USA

Serradilla MJ, Fotirić Akšić M, Manganaris GA, Ercisli S, González-Gómez D, Valero D (2017) Fruit chemistry, nutritional benefits, and social aspects of cherries. In: Quero-García J, Iezzoni A, Pulawska J, Lang G (eds) Cherries: botany, production and uses. CABI International, Boston, pp 420–441

Serrano M, Guillén F, Martínez-Romero D, Castillo S, Valero D (2005) Chemical constituents and antioxidant activity of sweet cherry at different ripening stages. J Agric Food Chem 53:2741–2745. https://doi.org/10.1021/jf0479160

Sharafi Y, Jannatizadeh A, Fard JR, Aghdam MS (2021) Melatonin treatment delays senescence and improves antioxidant potential of sweet cherry fruits during cold storage. Sci Hortic 288:110304. https://doi.org/10.1016/j.scienta.2021.110304

Shen X, Zhao K, Liu L, Zhang K, Yuan H, Liao X, Wang Q, Guo X, Li F, Li T (2014) A role for PacMYBA in ABA-regulated anthocyanin biosynthesis in red-colored sweet cherry cv. Hong Deng (*Prunus avium* L.). Plant Cell Physiol 55(5):862–880. https://doi.org/10.1093/pcp/pcu013

Shirasawa K, Isuzugawa K, Ikenaga M, Saito Y, Yamamoto T, Hirakawa H, Isobe S (2017) The genome sequence of sweet cherry (*Prunus avium*) for use in genomics-assisted breeding. DNA Res 24(5):499–508. https://doi.org/10.1093/dnares/dsx020

Singh JA, Bharat A, Edwards NL (2015) An internet survey of common treatments used by patients with gout including cherry extract and juice and other dietary supplements. J Clin Rheumatol 21:225–226. https://doi.org/10.1097/RHU.0000000000000246

Sooriyapathirana SS, Khan A, Sebolt AM, Wang D, Bushakra JM, Lin-Wang K, Allan AC, Gardiner SE, Chagné D, Iezzoni A (2010) QTL analysis and candidate gene mapping for skin and flesh color in sweet cherry fruit (*Prunus avium* L.). Tree Genet Genomes 6:821–832. https://doi.org/10.1007/s11295-010-0294-x

Starkevič P, Paukštytė J, Kazanavičiūtė V, Denkovskienė E, Stanys V, Bendokas V, Šikšnianas T, Ražanskienė A, Ražanskas R (2015) Expression and anthocyanin biosynthesis-modulating potential of sweet cherry (*Prunus avium* L.) *MYB10* and *bHLH* genes. PLoS One 10(5):e0126991. https://doi.org/10.1371/journal.pone.0126991

Sun L, Zhang M, Ren J, Qi JX, Zhang GJ, Leng P (2010) Reciprocity between abscisic acid and ethylene at the onset of berry ripening and after harvest. BMC Plant Biol 10:257. https://doi.org/10.1186/1471-2229-10-257

Teribia N, Tijero V, Munné-Bosch S (2016) Linking hormonal profiles with variations in sugar and anthocyanin contents during the natural development and ripening of sweet cherries. New Biotechnol 33:824–833. https://doi.org/10.1016/j.nbt.2016.07.015

Tsuda T (2012) Dietary anthocyanin-rich plants: biochemical basis and recent progress in health benefits studies. Mol Nutr Food Res 56:159–170. https://doi.org/10.1002/mnfr.201100526

Usenik V, Fabčič J, Štampar F (2008) Sugars, organic acids, phenolic composition and antioxidant activity of sweet cherry (*Prunus avium* L.). Food Chem 107:185–192. https://doi.org/10.1016/j.foodchem.2007.08.004

Usenik V, Fajt N, Mikulic-Petkovsek M, Slatnar A, Stampar F, Veberic R (2010) Sweet cherry pomological and biochemical characteristics influenced by rootstock. J Agric Food Chem 58:4928–4933. https://doi.org/10.1021/jf903755b

Vanderzande S, Zheng P, Cai L, Barac G, Gasic K, Main D, Iezzoni A, Main D, Peace C (2020) The cherry 6+9K SNP array: a cost-effective improvement to the cherry 6K SNP array for genetic studies. Sci Rep 10(1):7613. https://doi.org/10.1038/s41598-020-64438-x

Wang J, Liu W, Zhu D, Zhou X, Hong P, Zhao H, Tan Y, Chen X, Zong X, Xu L, Zhang L, Wei H, Liu Q (2020) A *de novo* assembly of the sweet cherry (*Prunus avium* cv. Tieton) genome using linked-read sequencing technology. PeerJ 5(8):e9114. https://doi.org/10.7717/peerj.9114e9114

Wei H, Chen X, Zong X, Shu H, Gao D, Liu Q (2015) Comparative transcriptome analysis of genes involved in anthocyanin biosynthesis in the red and yellow fruits of sweet cherry (*Prunus avium* L.). PLoS One 10(3):e0121164. https://doi.org/10.1371/journal.pone.0121164

Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol 126(2):485–493. https://doi.org/10.1104/pp.126.2.485

Wu T, Tang Q, Yu Z, Gao Z, Hu H, Chen W, Zheng X, Yu T (2014) Inhibitory effects of sweet cherry anthocyanins on the obesity development in C57BL/6 mice. Int J Food Sci Nutr 65:351–359. https://doi.org/10.3109/09637486.2013.854749

Yang H, Tian C, Ji S, Ni F, Fan X, Yang Y, Sun C, Gong H, Zhang A (2021a) Integrative analyses of metabolome and transcriptome reveals metabolomics variations and candidate genes involved in sweet cherry (*Prunusavium* L.) fruit quality during development and ripening. PLoS One 16(11):e0260004. https://doi.org/10.1371/journal.pone.0260004

Yang H, Tian C, Li X, Gong H, Zhang A (2021b) Transcriptome co-expression network analysis identifies key genes and regulators of sweet cherry anthocyanin biosynthesis. Horticulturae 7(6):123. https://doi.org/10.3390/horticulturae7060123

Yao H, Tian S (2005) Effects of pre- and postharvest application of salicylic acid or methyl jasmonate on inducing disease resistance of sweet cherry fruit in storage. Postharvest Biol Technol 35:253–262. https://doi.org/10.1016/j.postharvbio.2004.09.001

Zhang M, Yuan B, Leng P (2009) Cloning of 9-cis-epoxycarotenoid dioxygenase (NCED) gene and the role of ABA on fruit ripening. Plant Signal Behav 4(5):460–463. https://doi.org/10.4161/psb.4.5.8542

Zhang G, Sebolt AM, Sooriyapathirana SS, Wang D, Bink MCAM, Olmstead JW, Iezzoni A (2010) Fruit size QTL analysis of an F1 population derived from a cross between a domesticated sweet cherry cultivar and a wild forest sweet cherry. Tree Genet Genomes 6(1):25–36. https://doi.org/10.1007/s11295-009-0225-x

Zhao YY, Rosyara U, Iezzoni AF, Peace C, Whiting M, Dhingra A, Oraguzie N (2014) Identification of QTL underlying soluble solids content and titratable acidity in sweet cherry (*Prunus avium* L.). HortScience 49:149–150

Part V

Vegetable Crops

Potato Nutraceuticals: Genomics and Biotechnology for Bio-fortification

Teresa Docimo, Nunzia Scotti, Rachele Tamburino, Clizia Villano, Domenico Carputo, and Vincenzo D'Amelia

Contents

Abstract

Potato is a perfect target for bio-fortification strategies. This crop has contributed for thousand years to human diet, and its tubers still represent a staple food fundamental for worldwide food security. Besides being an important source of

T. Docimo · N. Scotti · R. Tamburino · V. D'Amelia (✉)
Institute of Biosciences and Bioresources, National Research Council of Italy (CNR-IBBR), Portici, Italy
e-mail: vincenzo.damelia@ibbr.cnr.it

C. Villano · D. Carputo
Department of Agricultural Sciences, University of Naples Federico II, Portici, Italy

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_48

energy, potato tubers contain compounds with nutraceutical properties including minerals, vitamins, proteins, and specialized metabolites. Therefore, we outlined this chapter devoting a significant space to both genomic and biotechnology studies addressed, respectively, to identify genomic sequence and genes influencing the amount nutraceutical molecules accumulated and to increment/introduce existing nutraceutical molecules or novel nutrients in the tuber. Before entering in the core of the chapter, we made an extensive introduction to the nutraceutical molecules which have been investigated in potato tubers. Then, we focused on innovative approaches, including gene editing and organelle transformation, with great potential for the nutraceutical bio-fortification of the potato. The *leitmotiv* of the chapter is the biodiversity of potato germplasm (including wild tuber-bearing species), which is a key aspect in potato tuber bio-fortification. The critical review of the literature suggested us to conclude pointing out the current scientific challenges for fully exploiting potato as a "vibrant feedstock for nutraceuticals."

Keywords

Tubers · Specialized metabolites · Vitamins · Minerals · Metabolic engineering · Genomic approach

1 Potato: Introduction to an Outstanding Food Crop for Nutraceuticals

The tubers of the cultivated potato (*Solanum tuberosum*) play a significant role in human nutrition and represent an established source of starch for the food industry. The worldwide production of potato stands 359 million tons, and its cultivation occupies 20 million hectares of farmland globally (FAOSTAT. In: FAO [online]. http://www.fao.org/faostat/). The biochemical composition of potato tubers varies according to the variety, agronomic practice, soil and climate, and storage condition. Because of the high concentration of carbohydrate in the form of starch (more than 60% of the dry matter), potato tubers are considered a rich energy providing food. One hundred grams of boiled tuber (that can correspond to a tuber of average size) provides about 110 Kcal, with less of 0.5 g of lipids (Burgos et al. 2020). However, potato is not only starch. A single tuber can also provide 50% of the recommended daily allowance of vitamin C and also balanced proteins of high biological value (i.e., the proportion of absorbed proteins which are incorporated into protein of the human body) (Burgos et al. 2020; Singh et al. 2020). The current research direction is toward valorizing the presence of important vitamins and phytochemicals (belonging to the primary and specialized metabolites) and exploiting potato proteins for several technological purposes (Hussain et al. 2021).

There are numerous valuable papers and book chapters which have analyzed or reviewed the nutraceuticals and biochemical composition of potato tubers (Burgos et al. 2020 and Singh et al. 2020 just to mention the most recent). All these works emphasized potato tubers as "vibrant feedstock for nutraceuticals" (to use the words

of Stewart and Taylor 2017), with a large biochemical variability enclosed in more than 10,000 varieties (Dolničar 2021). Polyphenols, vitamin C, vitamins of B group, and carotenoids have been particularly reviewed in these works which also underlined that many of these molecules are greatly accumulated in tuber peels, a typical by-product of food and starch industries. Another important aspect which is covered by studies dealing with potato nutraceuticals is the effect of cooking methods on these molecules (D'Amelia et al. 2022). Potatoes are not commonly consumed as fresh, and, therefore, the presence of toxic glycoalkaloids is one of the reasons why potato is needed to be cooked before consumption. Fortunately, these latter compounds are not generally stable at high temperature (D'Amelia et al. 2022); however, potato glycoalkaloids, which are especially accumulated in the peels, may even have positive function as anticancer and antiviral drugs (Hellmann et al. 2021).

Many of the minerals and molecules contained in potato tubers are indeed considered nutraceuticals, whose accumulation and bio-fortification can contribute to reach the nutritional and medical or health benefits in different countries, including the prevention and treatment of disease. Potato being the world's third most important crop in terms of human consumption, it is the perfect food crop to be enriched with functional molecules. Although the cultivated potato displays complex genetic features (a tetraploid – $2n = 4x = 48$ – nature, a high level of heterozygosity, a tetrasomic inheritance, and a severe inbreeding depression), there have been several genomic and metabolic engineering research studies aimed at the bio-fortification of its tubers with healthy molecules. Its metabolic and proteomic repertoire has also been exploited for different industrial applications.

The purpose of this chapter is to review these major researches and to identify novel cutting-edge technologies which can open new horizon of research in nutraceutical bio-fortification of potato. This first and second section of this chapter introduce to the metabolic and proteomic repertoire that characterize potato tubers from wild and cultivated species and to the genomic studies carried out to identify the regulation of nutraceutical molecules. The third section is the core of the chapter, where nuclear and organelle cis/transgenic along with next generation biotechnologies are reviewed. We concluded the chapter by providing critical perspectives on the exploitation of these technologies, underlining current challenges for potato tuber bio-fortification. We are not going to consider in detail aspects related to agronomic practices and conventional breeding approaches, which have been extensively described in the chapter of Bradshaw (2019), previously published in the book titled *Improving the Nutritional Value of Potatoes by Conventional Breeding and Genetic Modification*.

2 Nutraceutical, Metabolic, and Proteomic Repertoire in Potato Germplasm

Before the nineteenth century, potato was consumed mostly by Andes population, and then it was introduced to Europe to supply secure food for the large demographic growth. Today, population is expected to increase by two billion in 2050, and in view

of higher food demand, potato is again attracting great attention as food security crop since tubers supply nutrients at lower cost than other vegetables and fruits. The scientific communities are actively involved in maximizing potatoes' nutritional value.

From a nutraceutical point of view, potato skin is rich in dietary fibers, and the flesh provides carbohydrates and several vitamins such as vitamin C (ascorbic acid), vitamin B1 (thiamine), vitamin B3 (niacin), vitamin B6 (pyridoxine), vitamin B9 (folate), riboflavin, and pantothenic acid. Moreover, a relevant presence of minerals in the tubers, such as calcium (Ca), phosphorus (P), magnesium (Mg), sodium (Na), potassium (K), chloride (Cl), sulfur (S), and iron (Fe), concurs to the whole-body homeostasis maintenance. The nutritional composition of potato tubers is summarized in Fig. 1. Tables 1 and 2 report the dietary reference values according to EFSA (European Food Safety Authority) reports (EFSA NDA Panel (EFSA Panel on Dietetic Products 2015)). An adequate macro- and micronutrient uptake is necessary for human health; indeed, insufficient intake of vitamins and minerals can affect disease resistance, cognitive development, and physical growth and can lead to serious metabolic disorders and even death.

"Hidden hunger" refers to the deficiencies of multiple micronutrients in the diet, which is associated not only to poverty but also to bad dietary habits commonly due to the lack of diversification in food consumption and to diet rich in energy but poor in nutrients. By looking at the latter aspect, several programs are currently oriented toward fruit and vegetable bio-fortification, and this is also true for potato tubers (Ierna et al. 2020). Potato bio-fortification programs have been followed by the International Potato Center (CIP) for the past 15 years, by launching breeding

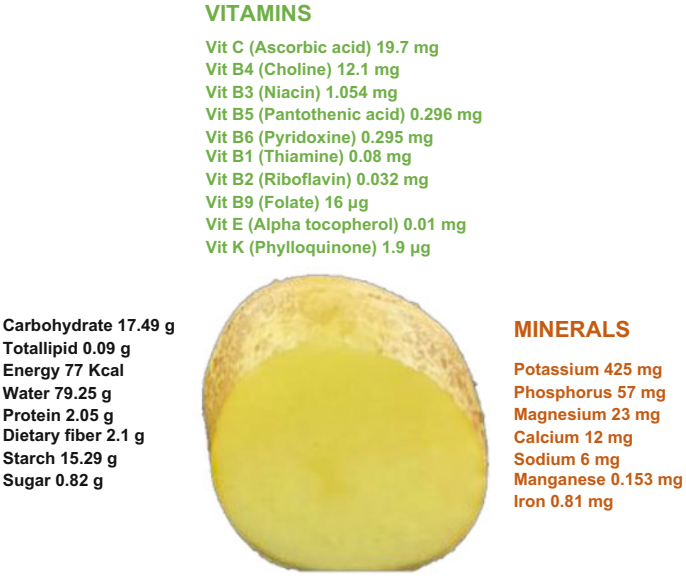


Fig. 1 Nutritional composition of potatoes per 100 g^{-1} FW

Table 1 Dietary reference values for vitamins and minerals for the EU according to https://multimedia.efsa.europa.eu/drvs/index.htm accessed on July 25, 2022

Nutrient (vitamins)	Gender (adults ≥ 18 years)	AI	AR	PRI	UL
Biotin	Both genders	40 μg/day	NA	NA	ND
Choline	Both genders	400 mg/day	NA	NA	NA
Cobalamin (vitamin B12)	Both genders	4 μg/day	NA	NA	ND
Folate	Both genders	NA	250 μg DFE/day	330 μg DFE/day	1000 μg/day
Niacin	Both genders	NA	1.3 mg NE/MJ	1.6 mg NE/MJ	900 mg/day nicotinamide
Niacin	Both genders	NA	1.3 mg NE/MJ	1.6 mg NE/MJ	10 mg/day nicotinic acid
Pantothenic acid	Both genders	5 mg/day	NA	NA	ND
Riboflavin	Both genders	NA	1.3 mg/day	1.6 mg/day	ND
Thiamin	Both genders	NA	0.072 mg/MJ	0.1 mg/MJ	ND
Vitamin A	Male	NA	570 μg RE/day	750 μg RE/day	3000 μg RE/day
Vitamin B6	Male	NA	1.5 mg/day	1.7 mg/day	25 mg/day
Vitamin B6	Female	NA	1.3 mg/day	1.6 mg/day	25 mg/day
Vitamin C	Male	NA	90 mg/day	110 mg/day	ND
Vitamin C	Female	NA	80 mg/day	95 mg/day	ND
Vitamin D	Both genders	15 μg/day	NA	NA	100 μg/day
Vitamin E	Male	13 mg/day	NA	NA	300 mg/day
Vitamin E	Female	11 mg/day	NA	NA	300 mg/day
Vitamin K as phylloquinone	Both genders	70 μg/day	NA	NA	ND

AI adequate intake, *AR* average requirement, *PRI* population reference intake, *UL* tolerable upper intake level

Table 2 Dietary reference values for minerals for EU according to https://multimedia.efsa.europa.eu/drvs/index.htm accessed on July 25, 2022

Nutrient (minerals)	Gender (adults ≥ 18 years)	AI	AR	PRI	UL	Safe and adequate intake
Calcium	Both genders 18–24 years	NA	860 mg/day	1000 mg/day	2500 mg/day	
Calcium	Both genders ≥ 25 years	NA	750 mg/day	950 mg/day	2500 mg/day	
Chloride	Both genders	NA	NA	NA	ND	3.1 g/day
Copper	Male	1.6 mg/day	NA	NA	5 mg/day	
Copper	Female	1.3 mg/day	NA	NA	5 mg/day	
Fluoride	Male	3.4 mg/day	NA	NA	7 mg/day	
Fluoride	Female	2.9 mg/day	NA	NA	7 mg/day	
Iodine	Both genders	150 µg/day	NA	NA	600 µg/day	
Iron	Male	NA	6 mg/day	11 mg/day	ND	
Magnesium	Male	350 mg/day	NA	NA	250 mg/day	
Magnesium	Female	300 mg/day	NA	NA	250 mg/day	
Manganese	Both genders	3 mg/day	NA	NA	ND	
Molybdenum	Both genders	65 µg/day	NA	NA	0.6 mg/day	
Phosphorus	Both genders	550 mg/day	NA	NA	ND	

Potassium	Both genders	3500 mg/day	NA	NA	ND	
Selenium	Both genders	70 μg/day	NA	NA	300 μg/day	
Sodium	Both genders	NA	NA	NA	ND	2 g/day
Zinc	Adults (LPI 300 mg/day)	NA	7.5 mg/day	9.4 mg/day	25 mg/day	
Zinc	Adults (LPI 600 mg/day)	NA	9.3 mg/day	11.7 mg/day	25 mg/day	
Zinc	Adults (LPI 900 mg/day)	NA	11 mg/day	14 mg/day	25 mg/day	
Zinc	Male adults (LPI 1200 mg/day)	NA	12.7 mg/day	16.3 mg/day	25 mg/day	
Zinc	Adults (LPI 300 mg/day)	NA	6.2 mg/day	7.5 mg/day	25 mg/day	
Zinc	Female adults (LPI 600 mg/day)	NA	7.6 mg/day	9.3 mg/day	25 mg/day	
Zinc	Adults (LPI 900 mg/day) Female	NA	8.9 mg/day	11 mg/day	25 mg/day	
Zinc	Adults (LPI 1200 mg/day)	NA	10.2 mg/day	12.7 mg/day	25 mg/day	

For adults, ARs and PRIs for zinc are provided for four levels of phytate intake (LPI): 300, 600, 900, and 1200 mg/day

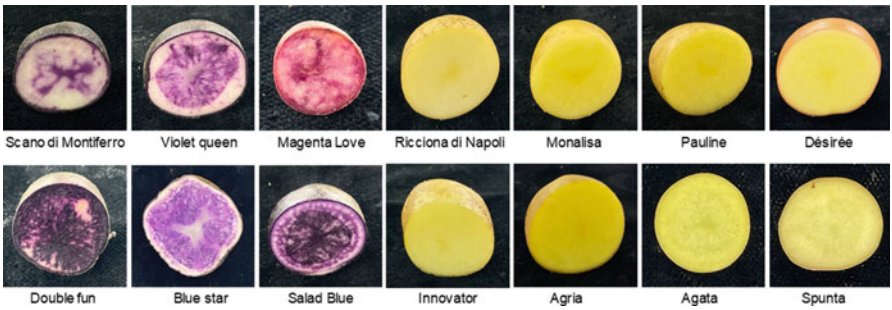


Fig. 2 Examples of commercial potato varieties and Italian landraces showing variable coloration of the flesh due to the different content of anthocyanins (purple/red) and carotenoids (yellow/orange)

programs oriented to increase the concentration of Fe and zinc (Zn), which raised almost of the doubled amount. However, the concentration of all these minerals and phytonutrients, besides being naturally different according to the varieties and agronomical practices, can be influenced by several other factors. For example, the amount of potato vitamins can be especially affected by pre- and postharvesting conditions. It should also be considered that the effective intake of potato nutraceuticals in the diet is largely influenced by their bioavailability. As matter of fact, cooking potatoes, without peeling off the skins, can guarantee a good intake of minerals and dietary fibers, but vitamins are largely reduced after cooking (Singh et al. 2020, 2022).

Carotenoids and polyphenols can be considered nutraceuticals, and these molecules are noteworthy present in potato tubers. Indeed, antioxidants in foods are key players against degenerative processes, cardiovascular diseases, and aging. Besides high-quality proteins, these antioxidants have probably been among the few nutraceuticals which domestication, breeding, and selection of potatoes have been addressed to. Carotenoids and anthocyanins characterize the color of potato tuber flesh. Commercially available, there are white, yellow, red, and purple fleshed potatoes (an example is reported in Fig. 2), whose flesh color intensity mirrors the amount of "dye" antioxidants. Yellow fleshed potatoes are richer in lutein and zeaxanthin carotenoids, whereas red and purple potatoes have a notable content in anthocyanins. These varieties can be ranged among the "functional foods."

2.1 Potato Nutrients: Starch, Sugars, and Lipids

Potato starch content is highly variable according to the genotype, environmental condition of growth, and agronomical practices. In general, fresh potatoes contain about 80% of water; 60–80% of the dry matter consists of starch. Potato starch is composed by amylopectin and amylose, with amylopectin ranging from 70% to 80% of the starch amount. The ratio between amylose (linear glucose polymers) and

amylopectin (highly branched glucose polymers) influences the degree of digestibility of starch, since amylose-rich starches are digested more slowly as opposed to starches with a high amylopectin content. The ratio *amylose/amylopectin* influences the glycemic index, which is a value indicating how much a specific food affects human blood sugar level (glucose). Starch is the major storage for carbohydrates, but potatoes have also free sugars mostly represented by the monosaccharide glucose and fructose and the disaccharide sucrose. Tubers of the tetraploid potato *S. tuberosum* have a total sugar amount up to a maximum of 300 mg 100 g^{-1} fresh weight (FW), while almost three times more (of about a maximum of 700 mg 100 g^{-1} FW) have been reported in tubers of diploid *S. phureja* (Duarte-Delgado et al. 2016). Potato sugar content is strongly dependent on the genotype. However, pre- and postharvest factors may play a crucial role. For example, particularly important are tuber maturity, growing temperature, agricultural conditions such as soil minerals and irrigation, mechanical stresses, and storage conditions. Sugar content in potato tubers is indeed particularly important since the high temperature of baking, frying, and roasting usually employed to cook potato, although are food safety measures, can also determine the formation of toxic molecules such as acrylamide. Acrylamide formation, which is considered a potential carcinogen, is present at elevated concentrations in different types of heat-treated foods, and it is formed during cooking at high temperature as reaction products of Maillard reaction between amino acids and reducing sugars. Therefore, a high basal level of potato sugars, as acrylamide precursors, and the specific processing conditions put potato in the group of food products with the highest level of acrylamide. In order to avoid acrylamide formation (probably carcinogenic to humans), the content of reducing sugar should be maintained below a level of 100 mg 100 g^{-1} FW (Liyanage et al. 2021).

Lipids are present in a relatively low amount, in a range of 0.1–0.5 g 100 g^{-1} FW. Phospholipids (47%), glycol, and galactolipids (22%) are the most represented. Their localization between the peel and the vascular ring of the tuber makes their content even reduced if the potatoes are thickly peeled. The composition in fatty acids of total lipids is mostly constituted of polyunsaturated fatty acids, e.g., linoleic and linolenic acids. Therefore, potato lipid composition is particularly "safe," since it is known that omega-3 fatty acids are beneficial against cardiovascular diseases and might also reduce the risk of type 2 diabetes.

Finally, potatoes contain fibers in the cell wall of the peel. These dietary components represent the portion of indigested carbohydrates which can be fermented in the large intestine. One hundred grams of potatoes cooked with the skin provide 2.1 g fiber, whereas for the same amount without skin the amount of fiber is reduced to 1.8 g. By way of comparison, potatoes contain less fiber than whole-grain cornmeal but more than white rice or whole-wheat cereal (1.6 g 100 g^{-1}).

2.2 Vitamins

Vitamins are molecules with different organic skeleton derived from various precursors. For example, vitamin C is a small soluble carbohydrate whose biosynthesis starts from glucose, and vitamin B complex belongs to pyridinecarboxylic acid

group of compounds. These molecules are not synthesized by humans, and their intake can occur only thought the diet. Deficiency of vitamin C causes scurvy, and the common symptoms of a poor vitamin C diet are fatigue, sore legs, and arms. The recommended dietary allowance (RDA) of vitamin C, as reported in Table 2, is about 90 mg daily for men and 80 mg for women and even more in particular condition such as pregnancy and for smokers. Potato, among vegetables, is a rich and inexpensive source of vitamin C, and its average content ranges from 84 to 145 mg 100 g^{-1} dry weight (Kawar 2016). Vitamin C content can vary according to the genotype, but the factor mostly influencing its content reduction is the oxidation. This can be caused by washing, peeling, slicing, cooking at high temperature, freezing, cooling, and storage (Sonar et al. 2020). Another important group of potato vitamins belongs to the B complex. Members of this group are vitamin B1 (thiamin), vitamin B2 (riboflavin), vitamin B3 (niacin or niacin amide), vitamin B5 (pantothenic acid), vitamin B6 (pyridoxine), vitamin B7 (biotin), vitamin B9 (folic acid), and vitamin B12 (cobalamins). The RDA for thiamin is 1.2 mg/day and 1.1 mg/day for adult man and woman, respectively, and potato varieties contain approximately 10 mg of thiamin in 100 g of tubers. Niacin can be assumed through the diet, but it can be also synthesized from the tryptophan. Its deficiency can cause pellagra, whose symptoms can be diarrhea, skin inflammation, and loss of memory (Fu et al. 2014); this disease is highly diffused among those populations relying on a maize based diet. It is estimated that potatoes can provide about one-tenth of adult's daily requirement of niacin considering that RDA values for niacin are 14 mg/day and 16 mg/day for men and women, respectively. Vitamin B5 and B6 are present in potato with similar amount, less than 1 mg for 100 g of potato, and RDA for vitamin B5 is close to 5 mg/day for both adult men and women and for vitamin B6 is about 2.0 mg/day and 1.6 mg/day for men and women. Both vitamin B5 and B6 vitamins are resistant to cooking and high temperatures. In this group of vitamins, vitamin B9 (folate) deserves particular attention because its deficiency is becoming a global dietary issue. Folic acid deficit promotes cell oxidation processes, and it causes DNA genomic instability. For this reason, increasing dietary intake of folate is particularly important during pregnancy, and folate based food supplements are recommended to reduce the risks of congenital disabilities (Wierzejska and Wojda 2020). Potato can provide about the 1–2% of the daily RDA for folic acid, which is about 0.2 mg.

Finally, K naphthoquinone is the last group of lipophilic antioxidant vitamins present in potato. Vitamin K has some essential beneficial properties deriving from its capability of blood clotting properties. Vitamin K is a group of chemically related compounds known as naphthoquinones and named vitamins K, K1, K2, and K3. One hundred grams of potatoes can contain from 1.5 to 3.0 micrograms of vitamin K, which is almost the 2–4% of the daily recommended value for an adult.

2.3 Specialized Metabolites

In industrialized countries, potato is erroneously associated to bad diet habits, mostly because of the consumption of french fries and chips. However, potato is recognized

by the Food and Agriculture Organization of the United Nations (FAO) as a staple and sustainable food for the growing world population (http://www.fao.org/potato-2008/en/aboutiyp/index.html, accessed on 2022). This is not only for the nutritional properties previously described but because it has a plethora of specialized metabolites whose consumption in the diet beneficially promotes human health. The high content of antioxidant compounds, mainly carotenoids and phenylpropanoids, makes tubers an optimal functional food. Carotenoids in plants are more than 700; they are isoprenoid-based compounds typically represented by a tetraterpene skeleton. The modification of this backbone influences color and antioxidant activities of tubers. Carotenoid content ranges from 2.7 to 7.4 μg g^{-1} FW in commercial white potatoes. The highest amounts of carotenoids are present in yellow- and orange-fleshed potatoes. Zeaxanthin is the carotenoid most responsible for orange color, whereas lutein is responsible for yellow. Andean cultivated landraces have been reported to accumulate a very high content of carotenoids; the concentration of lutein/zeaxanthin may reach up to 18 μg g^{-1} DW and about 2 μg g^{-1} DW of β-carotene (Milczarek and Tatarowska 2018). Similarly, diploid *S. stenotomum* and *S. phureja* may contain up to 20 μg g^{-1} FW of zeaxanthin (Brown et al. 2019). Through metabolic engineering approaches, β-carotene content in Désirée was increased in up to about 50 μg g^{-1} DW (Diretto et al. 2007a). Although this strategy resulted very efficient to produce what is referred as "golden potato," its consumption is limited to few European and American countries due to specific regulatory limitations and also consumer acceptance.

Among antioxidants, hydroxycinnamic acids are the most abundant with chlorogenic acid (CGA) as the most accumulated. In white and yellow potatoes, CGA can constitute 90% of tuber total soluble phenolics, with 5-*O*-caffeoylquinic acid being the most abundant CGA. Two hundred grams of potatoes could provide over 250 mg of CGA, which is much more than the amount found in a cup of coffee (the most known source of CGA) (Lu et al. 2020). Interestingly, red and purple potatoes produce an even higher quantity of CGA than white-fleshed tuber. Tubers also contain flavonoids, including anthocyanins and flavonols. Flavonoids have a C6-C3-C6 backbone; acylation, hydroxylation, methylation, and glycosylation of this skeleton give rise to thousands of compounds. Anthocyanins are synthesized in the general flavonoid pathway, and the first committed step in the anthocyanin pathway is catalyzed by dihydroflavonol reductase (DFR). The conversion from colorless leucoanthocyanidins to anthocyanidins by anthocyanidin synthase is the coloring reaction of the biosynthetic pathway. The six major anthocyanidins in potato are cyanidin, delphinidin, malvidin, pelargonidin, peonidin, and petunidin; the degree of hydroxylation/methoxylation of the B-ring influences the color of the tuber. For example, B-ring hydroxylation increases blue color, whereas methylation can turn the tuber red. Higher hydroxylation on the B-ring positively influences the antioxidant activity.

2.4 Glycoalkaloids

Another class of important and characteristic solanaceous specialized metabolites is glycoalkaloids. These molecules, due to their toxicity, play an important defensive

role in plants against pathogens and predators (Cárdenas et al. 2019). Main glycoalkaloids in cultivated potato are α-solanine and α-chaconine, where the latter is usually present in higher amounts than solanine (D'Amelia et al. 2022). These molecules consist of a lipophilic steroid nucleus, which is extended by two fused nitrogen-containing heterocyclic rings at one end and bound to a polar water-soluble trisaccharide at the other. The level of toxic glycoalkaloids in potatoes varies depending on the variety, growing conditions, storage and transportation, temperature, cutting, sprouting, and general exposure to any biotic or abiotic stress. Glycoalkaloids are present at high concentration in the skin of tubers, and higher concentrations are found around the potato eyes and wounded areas and in the sprouts. On average, a potato tuber can contain about 12–20 mg kg^{-1} of glycoalkaloids, while a green and germinated tuber reaches 250–280 mg kg^{-1}. It is worth to mention that a toxic dose has been estimated at approximately 2–5 mg kg^{-1} body weight, whereas a lethal one is around 3–6 mg kg^{-1} of body weight (Schrenk et al. 2020).

2.5 Potato Proteins and Peptides

Potato protein content ranges from 1% to 1.5% of tuber fresh weight. Despite the low content in tubers, these proteins have high nutritional value because they are rich in essential amino acids such as lysin, threonine, tryptophan, and methionine (Kärenlampi and White 2009). Therefore, they are nutritionally superior to those of many other plants (Kärenlampi and White 2009).

Potato proteins are obtained as by-products of the starch industry and are commonly divided into three groups: (i) patatin, (ii) protease inhibitors, and (iii) high-molecular-weight proteins (Kowalczewski et al. 2019). Patatin, also known as tuberin, is a 40–45 kDa glycoprotein that represents 40% of total soluble proteins (TSP) in potato tubers (Lehesranta et al. 2005). It occurs in multiple isoforms being encoded by two multigene families, and different potato varieties show distinctive patterns of putative patatin isoforms. As an example, Désirée showed nine isoforms, whereas Bintje four (Lehesranta et al. 2005). Protease inhibitors (PIs) also account for 40% of TSP and are the structurally heterogeneous group with broad range of antifungal and antimicrobial activities (Bártová et al. 2019). PIs are classified into seven different families, based on their target, structural properties, and mechanism of action: protease inhibitors I (PI-1) and II (PI-2), potato cysteine protease inhibitors (PCPI), potato aspartate protease inhibitors (PAPI), potato Kunitz-type protease inhibitor (PKPI), potato carboxypeptidase inhibitors (PCI), and other serine potato protease inhibitors (Meenu Krishnan and Murugan 2016). The third group of potato proteins consists of high-molecular-weight proteins mainly involved in starch biosynthesis that have not been fully studied.

Potato proteins are gaining increasing interest in food industry applications for their well-known functional features. These include activities that are strictly linked to food processing such as emulsifying, foaming, gel forming, antifungal, antimicrobial, and antiviral properties. They also show multiple health promoting

properties. Particularly, patatin demonstrated antioxidant and lipase activities (Wu et al. 2021). These have been associated to protection against oxidative stress in chronic degenerative disease, anticancer activity, and anti-obesity effects (Wu et al. 2021). Similarly, protease inhibitors showed anticarcinogenic activity through three main mechanisms: interfering with tumor-cell proliferation, forming hydrogen peroxide, and blocking solar ultraviolet (UV) irradiation (Sitjà-Arnau et al. 2005). Potato protease inhibitors also increased the level of the peptide cholecystokinin that increases satiety via trypsin inhibition, thus reducing food intake in humans (Komarnytsky et al. 2011). Other studies have demonstrated that peptides obtained from potato protease inhibitors positively affect serum lipids (e.g., cholesterol) through their sterol binding properties (Liyanage et al. 2010). Tuber proteins can be hydrolysate (e.g., Alcalase treatment) to produce bioactive peptides with multiple and enhanced pharmacological functions. Potato protein hydrolysate (PPH) revealed several beneficial effects: antihypertensive, through the enhanced inhibition of the angiotensin converting enzyme I (ACE); protective against hepatic and cardiac functions, through lipolytic activity; and anti-inflammatory by the activation of AMPK, a major signaling molecule controlling the pathways of hepatic metabolic homeostasis (Boudaba et al. 2018). High-quality, allergy-free, and clear activity against various disease conditions described supports the promising usage of potato proteins and peptides in novel therapeutic food products.

3 A Glance on Potato Conventional Breeding: Mapping of Gene/QTLs

As previously stated, conventional breeding approaches, addressed at improving the nutritional value of potato tubers, have been largely reported by the recent work of Bradshaw (2019). Here, we have summarized few aspects related to the limit of conventional breeding and the mapping of QTLs for potato, aspects which extend also to the bio-fortification.

Most crops are polyploids and potato is among these ensuring fulfilling food demand. Due to its complex genetics, potato breeding is a heavy task. The introduction of favorable traits in potato by means of conventional breeding techniques is very laborious and time consuming. The polyploidy offers several advantages such as heterosis and gene redundancy but also produces high heterozygosity which translates in broad segregation for many traits during breeding (Bonierbale et al. 2020). Cultivated potatoes *Solanum tuberosum* (ssp. *tuberosum*) is an autotetraploid ($2n = 4x = 48$), and four homologous chromosomes are matched pairs during meiosis process. This genetic conformation results in a tetrasomic inheritance which complicates the studying of locus interactions, allelic combinations, and genetic effects, and these effects become even more complicated for quantitative (polygenic) traits such as the amount of nutritional compounds and proteins. In this context, the pyramiding of desirable alleles in a potato breeding line or population requires screening of a progeny of thousands or even million seedlings for the identification of the desirable traits into a single clonal line, thus discouraging

progress in breeding improvements for tetraploid potatoes. In the last 10 years of the last century, various molecular markers were made available such as random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), inter simple sequence repeats (ISSR), inter-retrotransposon amplified polymorphism (IRAP), and simple sequence repeats (SSR), and lately high-throughput assays, such as single-nucleotide polymorphism (SNP) markers, have become increasingly important in genetic studies for several crops. However, in autotetraploid genetic mapping technique, such as linkage and quantitative trait locus (QTL) analysis, the genomic complexity can limit assessing correctly allelic dosage (difficulty to determine the correct allelic combination and frequency of recombination between loci), thus jeopardizing proper linkage map construction and quantitative trait loci (QTL) mapping (Bradshaw 2019). In this regard, for SNP array in polyploids, two important platforms, Illumina Infinium and Affymetrix Axiom, have been developed. The development of a potato array provided a genome-wide set of single-nucleotide polymorphism (SNP) markers which, along with the development of statistical models and suitable software, have significantly increased the power of SNP data (Hackett et al. 2017). Further, the development of Tetraploid SNP Map, a user-friendly software specifically designed to analyze SNP markers in polyploid germplasm (Hackett et al. 2017), encouraged implementation of QTL analysis with tetraploid mapping populations of potato. At present, the hybridization based SNP array and next generation sequencing (NGS) enabled the use of genotyping by sequencing (GBS) as the most popular high-throughput genotyping platforms.

Table 3 reported a summary of QTL studies related to quality traits in potato carried on biparental population, selected for their allelic variation affecting the trait of interest.

Table 3 Summary of QTL mapping studies related to quality traits of potato

Traits	Markers	QTL
Tuber starch content (TSC)	DArT RFLP	QTL on chr 1
Tuber starch and tuber starch content	AFLP and SSR	QTLs on chr 5, 8, 9, 10, 12
Tuber starch	RFLP, AFLP	QTLs on all chromosomes
Dry matter content (DMC)	AFLP and SSR, SNPs	QTLs on chr 2, 3, 5, and 7, 8, and 11
Specific gravity (SG)	RFLPs, RAPDs, AFLP, and SSR	QTLs on chr 1, 2, 3, 5, 7, and 11
Cold induced sweetening	AFLP and SSR	QTLs on chr 5 and 1 on chr 8
Amylose content	AFLP and SSR	QTLs at chr 2, 3, 5, 7, and 10
Sugars	RFLP, AFLP	QTLs on all chromosomes
Glucose and fry color	SolCAP 8 K SNP array	QTLs on chr 4, 5, 6, 10, and 11
Skin texture	AFLP and SSR	QTL on chr 2, 6, and 12
Tuber flesh color	AFLP, CAPS, SSR	QTLs on chr 5, 8, and 9
Tuber carotenoids	AFLP, SSR, and DarT	QTLs on chr 3 and 9

Modified from Naeem et al. (2021) and reference therein

Although several candidate genes and QTLs were identified, the detected QTLs represent small haplotype blocks since it is restricted only to the diversity of the two parents. To obtain more stable and reliable QTL, it is necessary to enhance mapping resolution.

Currently, efforts have been started to produce diploid potato parental lines for breeding. The use of diploid based inbreed lines may open potato breeding to more sophisticated mapping population such as recombinant inbred lines (RILs), introgression lines (ILs), and multi-parent advanced generation intercross (MAGICs) populations to identify gene-trait association. In this perspective, the work of Zhang and collaborators published on *Cell* is a smart example of how to develop diploid potato inbred lines and vigorous F1 hybrids by using genome design approach. Surely, this approach could be also used in the future to remodel the potato nutraceutical value.

4 An Overview of the Genomic/Transcriptomic Strategies to Help Nutraceutical Bio-Fortification in Potato Tubers

The sequencing of the potato genome, together with the successive re-annotations (http://spuddb.uga.edu), allowed scientists to carry out several transcriptomic (e.g., microarray, RNA sequencing) and genomic studies (e.g., single-nucleotide polymorphism (SNP) genotyping, genome-wide association studies (GWAS), comparative genomics) that have been strategic for examining the genetic diversity of potato germplasm (among recent papers, see Tang et al. 2022; Zhang et al. 2022). The sequencing of a vast number of plant genomes allowed to perform comparative analyses of nucleotide sequences to evidence similarities and diversity between related species and uncover evolutionary relationships. Wild potatoes, for example, have been used in the past to introgress novel traits, and related genes, in the cultivated potatoes. However, these species have been used to a limited extent for improving the nutritional value of cultivated potatoes (Bradshaw 2019). It is only in recent years that breeders have considered the increment of nutraceuticals as target trait in breeding programs (e.g., https://potatoes.colostate.edu/potato-breeding/; https://potato.tamu.edu/program/); in this scenario, genomic resources provide an important contribution in identifying genes involved in the accumulation of nutraceutical molecules. As genome sequencing has revolutionized all research approaches, in this paragraph we are going to concentrate to those studies that exploited the sequence of the potato genome to identify genes (or classes of genes) with a direct/indirect effect on the accumulation in the tubers of compounds recognized as nutraceuticals.

4.1 Mineral Elements

Mineral elements are important supplements in human diet. Variability in the amounts of mineral elements is present in potato germplasm. Differences are present

between modern varieties especially for N, P, K, S, Ca, Mg, Cu, Fe, Mn, and Zn; between Andean landraces as regards Ca, Fe, and Zn; and between tuber-bearing wild species as regards Ca (Bradshaw 2019). Though the amount of mineral elements that accumulates in potato tubers showed good heritability, bio-fortification through breeding is not sufficient to increase the concentration of certain elements, and it is needed to be complemented with fertilization strategies (Kromann et al. 2017). Moreover, with the aim to increase the content of mineral elements, it should be taken into account that a complex network of homeostasis mechanism guarantees an adequate and not toxic level of mineral elements and that an important regulation between source and sink is present. For example, Mengist and colleagues (Mengist et al. 2018) showed that the accumulation of Zn in tubers is regulated by stems and also identify a QTL controlling the correlation between Zn and Cd, the latter being toxic to humans (Mengist et al. 2018). A reverse genomic approach can be an alternative to resolve the complex network governing mineral uptake. For example, by analyzing on nucleotide sequences which encode for transporters with important function on mineral accumulation, it could be possible to identify allelic variants with a higher efficiency in increasing the mineral element concentration in tubers or to be used as candidate gene in biotechnological approaches. Twenty-three putative transporters with specific affinity for K (belonging to the family of High-Affinity Potassium transporter (HKT), Potassium Proton Antiporters (KEAs) and Proton-coupled potassium transporters (KUP/HAK/KTs), 8 genes encoding for high-affinity transporter of phosphorus (PHT1) family (Cao et al. 2020; Azeem et al. 2021), and 12 SULTR involved in sulfur transportation (Vatansever et al. 2016) have been identified in potato. Transporters with wider affinity for minerals have been also characterized. For example, 36 gene members of heavy metal ATPase (HMA) family and 21 of zinc/iron-regulated transporter-like (ZIP) gene family were systematically analyzed (He et al. 2020; Li et al. 2020). Aquaporins are also important proteins involved in the transcellular membrane transport of water and other small solutes like silicon (Si). For these transporters, 41 encoding genes have been found within the potato genome (Venkatesh et al. 2013).

4.2 Vitamins

Vitamin C, B6, and folate (vitamins B9) are the most represented vitamins in the potato tubers. These vitamins showed a very diversified concentration among varieties and wild species. Several have been the breeding efforts addressed to increase the content of vitamin C. Though the first evidence for the genetic basis of vitamin C has been reported more than 30 year ago and studies are still coming along showing the differences of concentration of this vitamin within potato germplasms, the literature regarding the identification of major loci or marker affecting vitamin C accumulation specifically in tubers is quite rare. However, there is a great interest to increase the accumulation of vitamin C in the tubers of potato, and substantial efforts have been made through biotechnological approaches

(described later in this chapter) involving genes of the vitamin C pathway (Goo et al. 2008; Hameed et al. 2018). As regards other vitamins, the literature reviewed indicates interest of plant scientists to use genomic tools to identify loci involved in their accumulation. For example, Bali et al. (2018) identified a total of 497 significant SNPs associated with folate (vitamin B9) in a diploid segregating population developed from *S. boliviense*. Eighteen of these markers were directly associated with folate metabolism. The sequenced potato genome allowed also the identification of genes encoding for proteins forming the multimeric complex necessary for vitamin B6 biosynthesis. These proteins were identified in a potato cDNA library by using an *Arabidopsis thaliana* pyridoxal biosynthesis (PDX) protein as bait in a *yeast two hybrid* screening (Mooney et al. 2013). However, no studies have been carried out so far to understand differences at genomic level that are responsible for the different accumulation of vitamin B6 in tubers.

4.3 Specialized Metabolites

In the last 15 years, a high number of researches have been dedicated to identify genomic regions affecting or associated with the accumulation of carotenoids and phenylpropanoids in potato. Specific markers associated to accumulation of carotenoids tubers of potatoes have been developed by using diploid populations. SNP analyses in both diploid and tetraploid populations marked one dominant allele of beta-carotene hydroxylase 2 (*CHY2*) (Y locus on chr3) having a major effect in changing white into yellow flesh color. *CHY2,* when combined with a recessive *Zeaxanthin epoxidase* (*zep*) allele, produced orange-fleshed tubers that accumulated large amounts of zeaxanthin (Sulli et al. 2017). In another diploid population, genome-wide QTL mapping combined with expression QTL (eQTL) analyses was used to identify another major carotenoid locus on chr9 and a potential gene candidate annotated as *early light inducible* protein (Campbell et al. 2014). Phenylpropanoids represent probably the class of healthy compounds mostly studied in potato tubers. In particular, anthocyanins and chlorogenic acid followed by hydroxycinnamic acids and flavonols have been subjected to different genic, molecular, genomic, or transcriptomic researches (e.g., D'Amelia et al. 2018; Bao et al. 2022). Being anthocyanins visible at naked eyes, earlier genetic investigations were particularly focused on the inheritance of loci influencing the presence or absence of red and purple pigmentation of tuber skin and flesh (De Jong et al. 2004). Anthocyanin regulatory elements (mainly characterized by the complex of R2R3 MYB, bHLH, and WD40 transcription factors) activate the genes encoding for relevant enzymes of the pathway. In potato, genes having these roles have been largely characterized by genetic, transcriptomic, and genomic approaches. The utility of RNA data resides in the fact that they evidenced the structural genes differently expressed between fleshes of pigmented and white tubers or between purple- and red-fleshed tubers. Consequently, lists of genes mainly involved in anthocyanin accumulation or in anthocyanin post-biosynthetic modifications (decorations of molecules to produce the different anthocyanin fractions) have been obtained (Ahn et al. 2022).

Compelling genomic outcomes are those which revealed noncoding RNAs showing a regulatory activity on the expression of both regulatory and structural genes. For example, a microRNA (miR828) was identified to influence the biosynthesis of anthocyanins in tuber flesh by targeting a R2R3 MYB transcription factors (Bonar et al. 2018). Another work, though carried out by using leaf materials, revealed potential long noncoding RNA (lncRNA) regulating the expression of genes encoding for important enzymes of phenylpropanoid pathway (i.e., phenylalanine ammonia lyase, flavanone 3-hydroxylase, and chalcone synthase) (Bao et al. 2022). Chlorogenic acids largely contribute to the antioxidant repertoire of potato tubers. However, genomic results on this compound came out mainly from studies conducted for investigating anthocyanins. A strong metabolic flux connection is present between anthocyanins and chlorogenic acids, and a molecular co-regulation is also hypothesized as side of anthocyanin works (Rommens et al. 2008). In recent work of Yang et al. (2021), a natural variation of the major isoform of chlorogenic acid was screened in diploid populations including 40 wild species and 374 landraces. The authors revealed the presence of 18 SNPs associated with this compound. Considering that consumers prefer white-fleshed varieties of potato rather than colored flesh (Barlagne et al. 2021), the possibility to increase the presence of this important antioxidant also in not purple/red potatoes is attractive.

4.4 Protein Content

Hereditability of total protein content is moderate, and it is a particularly complex regulated trait influenced by inter-locus interactions and environmental factors (Klaassen et al. 2019a, b). Genome sequencing coupled with transcriptomic and proteomic studies allowed to dissect the genetic architecture influencing this trait and to identify loci that can be used in breeding or biotechnology programs. Main outputs of these studies have been the identification of QTL, eQTL, and proteomic QTL (pQTL). Acharjee et al. (2018) used diploid potatoes and a proteomic approach to map pQTLs for over 300 different protein spots on every chromosome. Klaassen et al. (2019a) mapped QTLs on chromosomes 1, 3, and 5 using a biparental tetraploid population genotyped through SNPs. The QTL on chromosome 5 was then confirmed by a study which used a panel of tetraploid varieties in a GWAS (Klaassen et al. 2019b).

5 Metabolic Engineering for Nutraceuticals

Potato is among the most widely consumed vegetable in the world. Because of it is adaptability to different environments and its yielding capacity, potato plays an important role in worldwide food security. Potato tubers are rich in carbohydrates, high-value proteins, vitamins (vitamin c, above all), and also antioxidant molecules (e.g., anthocyanins and chlorogenic acid). However, the concentration of most of these compounds is often not sufficient, especially after cooking treatment, to satisfy the RDA for those populations where a diversity of diet is limited. Metabolic engineering

approaches accelerate the bio-fortification of crop with nutraceuticals compared to conventional breeding process or when natural variation in sexually compatible germplasm is insufficient to achieve satisfactory level (Van Der Straeten et al. 2020).

In this paragraph, examples of genetic engineering approaches to accelerate bio-fortification will be described. With the aim of meeting different readers' expertise, we will introduce various concepts related to the genetic transformation (trans−/intra−/cisgenesis), giving also a general idea of the techniques. We will also discuss non-conventional methods (such as organelle transformation) and gene editing as alternative or emerging techniques.

5.1 Conventional Genetic Transformation

Genetic transformation consists in the targeted manipulation of specific plant characters using genes from a range of sources. Through genetic engineering techniques (named transformation techniques), foreign genetic material can be added within a plant genome. Nowadays, several transformation techniques have been tuned in potato. The most used are the DNA uptake into isolated protoplasts mediated by chemical procedures and the *Agrobacterium*-mediated transformation in leaf explants. The frequency of stable transformants is higher when the *Agrobacterium*-mediated transformation is used. It is, indeed, the most commonly used method for plant genetic engineering. The pathogenic soil bacteria *Agrobacterium tumefaciens* that cause crown gall disease have the ability to introduce a tumor-inducing (Ti) plasmid into the nuclear genome of infected plant cells. Biotechnologists were able to engineer the Ti plasmid introducing the gene of interest within the transfer DNA or tDNA containing a regulatory sequence (including the promoters), the gene of interest, and the selection marker gene.

Various protocols based on *Agrobacterium*-mediated transformation available in potato and transgenic potato varieties have been approved for planting and commercialization. Potato GM varieties, approved for public uses, are listed in The International Service for the Acquisition of Agri-biotech Applications database at the Internet website www.isaaa.org. In most cases, these varieties have been withdrawn from the market because of low sales due to consumer/farmer rejection (Mi et al. 2015). For this reason, during the last decades, alternative methodologies of genetic transformation, such as cisgenesis and intragenesis, have been set up in metabolic engineering. Cis- and intragenesis are based on the exclusive use of genetic material from the same species or genetic material from closely related species capable of sexual hybridization. This is in contrast to transgenesis where one or more genes and DNA sequences can derive either from any organism or from a sexually incompatible donor plant. In detail, intragenesis and cisgenesis is a genetic transformation where genes are transferred from the same plant or a close relative plant species and any transgene, such as a selection marker gene and vector-backbone genes, is absent in the progeny. The main difference between intragenesis and cisgenesis is based on the concept that in the former methodology, the combination of genetic element through in vitro rearrangement is allowed. The topic of the use of metabolic

engineering approaches to increase nutrients in potato tubers has been reported by (Bradshaw 2019); more recently, several metabolic engineering approaches for the enhancement of potato nutritional value have been reported (Upadhyaya and Bagri 2021). In this chapter, a number of representative transgenic and cisgenic approaches have been itemized in Table 4, detailing the origin of the gene of interest and the potato variety used for the transformation. To obtain a successful bio-fortification of potato tubers by genetic transformation, it is fundamental to select the correct promoters. Indeed, this will drive a correct spatial and temporal expression of the selected gene/s. Therefore, in Table 4, we also summarized the promoters chosen by previous researchers in potato genetic transformation experiments.

Numerous studies have been addressed to increase the level of several types of carotenoids in the flesh of potato tubers, and many studies were particularly interested in the β-carotene, the precursor of vitamin A (Table 4). The earliest strategies attempted to modulate the flux of precursors through the introduction of either genes encoding for enzymes of the primary metabolism or belonging to early carotenoid biosynthetic steps. An important increment (sixfold) in β-carotene was obtained by introducing the gene encoding for the bacterial phytoene synthase (*CrtB*) under the control of the tuber-specific B33 promoter (Ducreux et al. 2005). However, the most successful approach was accomplished by the coordinated expression of genes isolated from *Erwinia herbicola.* In particular, under a tuber-specific promoter, genes encoding for the CrtB, the phytoene dehydrogenase (CrtI), and the lycopene β-cyclase (*CrtY*) were introduced (Diretto et al. 2007a). The authors developed the "golden tubers," where the provitamin A β-carotene increased by 3600-fold (about 47 mg g^{-1} of dried weight). More recently, Campbell et al. (2015) combined β-carotene ketolase (CrtW) and β-carotene hydroxylase (CrtZ) from bacteria with the *Orange* (*Or*) gene from cauliflower leading to substantial increment of the concentration of astaxanthin and also total ketocarotenoids. *Or* was identified from the spontaneous cauliflower orange mutant which accumulates high level of β-carotene in chromoplasts (Lu et al. 2007). Orange protein regulates carotenoid accumulation by posttranscriptionally regulating phytoene synthase, promoting the formation of chromoplast, and also preventing process of carotenoid degradation (Osorio 2019). The cauliflower *Or* was also overexpressed alone in potato, and the obtained transgenic tubers showed chromoplast neoformation and more than tenfold as much β-carotene as the level of non-transgenic cold-stored tubers (Lopez et al. 2008). Overall, these latter studies suggest that there is still a large possibility to further increase the total carotenoids or specific molecules in potato by combining the expression of biosynthetic genes and their regulators. Several regulatory genes of carotenoid pathway have been characterized in tomato (e.g., D'Amelia et al. 2019; Tang et al. 2022). Considering the phylogenetic relatedness between the two Solanaceae species, it is likely to have some success in the increment of carotenoid content by transgenically expressing these tomato genes in potato tubers.

Vitamin E and vitamin B6 have been also successfully increased through transgenic approaches (Table 4) (Crowell et al. 2008; Bagri et al. 2018). The strategy used by Qin et al. (2011) and De Lepeleire et al. (2018) to increase the content of vitamin C and folate (vitamin B), respectively, is quite interesting to discuss. The former can

Table 4 Examples of genetic transformation used for nutraceutical bio-fortification of potato tubers

Compound	Gene/s and origin	Genotype	Promoter	Methodology/ effect	References
Carotenoid	*CrtB* from *Erwinia uredovora*	*S. tuberosum* cv. Désirée and *S. phureja* cv. Mayan Gold	B33 patatin	Transgenesis/ activation	Ducreux et al. (2005)
Carotenoid	*CrtW*/CrtZ from *Brevundimonas* sp. SD212	*S. phureja* cv. Mayan Gold	GBSS	Transgenesis/ activation	Campbell et al. (2015)
Carotenoid	*Orange* gene from *B. oleracea*	*S. tuberosum* cv. Désirée	GBSS	Transgenesis/ activation	Lopez et al. (2008)
Carotenoid	*LCY-E* from *S. tuberosum* cv. Désirée	*S. tuberosum* cv. Désirée	B33 patatin	Cisgenesis/ silencing	Diretto et al. (2006)
Carotenoid	*CrtB*, *CrtI*, and *CrtY* from *E. herbicola*	*S. tuberosum* cv. Désirée	CaMV 35S	Transgenesis/ activation	Diretto et al. (2007a)
Carotenoid	*CrtB*, *CrtI*, and *CrtY* from *E. herbicola*	*S. tuberosum* cv. Désirée	B33 patatin	Transgenesis/ activation	Diretto et al. (2007a)
Carotenoid	*CHY1/CHY2* from *S. tuberosum* cv. Désirée	*S. tuberosum* cv. Désirée	B33 patatin	Cisgenesis/ silencing	Diretto et al. (2007b)
Vitamin E	*HPPD/HPT* from *A. thaliana*	*S. tuberosum* cv. Spunta and MSE149-5Y	CaMV 35S	Transgenesis/ activation	Crowell et al. (2008)
Folate	*GTPCHI/ADCS* from *A. thaliana*	*S. tuberosum* cv. Désirée	GBSS	Transgenesis/ activation	Blancquaert et al. (2013)
Folate	*GTPCHI/ADCS* from *A. thaliana*	*S. tuberosum* cv. Désirée	B33 patatin	Transgenesis/ activation	Blancquaert et al. (2013)
Folate	*FPGS* from *A. thaliana* and *HPPK/DHPS* from *O. sativa*	*S. tuberosum* cv. Désirée	B33 patatin	Transgenesis/ activation	De Lepeleire et al. (2018)
Ascorbic acid	*DHAR* from *S. indicum*	*S. tuberosum* cv. Jowon	B33 patatin	Transgenesis/ activation	Goo et al. (2008)
Ascorbic acid	*DHAR* from *S. indicum*	*S. tuberosum* cv. Jowon	CaMV 35S	Transgenesis/ activation	Goo et al. (2008)

(continued)

Table 4 (continued)

Compound	Gene/s and origin	Genotype	Promoter	Methodology/ effect	References
Ascorbic acid	*GLOase* gene rat	*S. tuberosum* cv. Taedong Valley	CaMV 35S	Transgenesis/ activation	Upadhyaya et al. (2010)
Ascorbic acid	*DHAR* gene from *S. tuberosum*	*S. tuberosum* cv. Favorita	CaMV 35S	Cisgenesis/ activation	Qin et al. (2011)
Vitamin B6	*PDXII* from *A. thaliana*	*S. tuberosum* cv. Kufri chipsona	CaMV 35S	Transgenesis/ activation	Bagri et al. (2018)
Triacylglycerol	*DGAT1* from *A. thaliana* and *WRI1*, *OLEOSIN* from *S. indicum*	*S. tuberosum* cv. Atlantic	CaMV 35S/B33 patatin	Transgenesis/ activation	Liu et al. (2016)
Calcium	*CAX1* from *A. thaliana*	*Solanum tuberosum* var Russet Norkotah	CaMV 35S	Transgenesis/ activation	Park et al. (2005)
Calcium	*CAX1* from *A. thaliana*	*Solanum tuberosum* var Russet Norkotah	CDc2a	Transgenesis/ activation	Park et al. (2005)
Calcium	*CAX2* from *A. thaliana*	*Solanum tuberosum* var Daejiree	CaMV 35S	Transgenesis/ activation	Kim et al. (2006)
Lipid	*9-LOX*	*S. tuberosum* cv. Désirée	CaMV 35S	Cisgenesis/ silencing	Göbel et al. (2003)
Lipid	*9-LOX*	*S. tuberosum* cv. Désirée	CaMV 35S	Cisgenesis/ silencing	Eschen-Lippold et al. (2007)

be considered a cisgenic approach because the authors activated the gene encoding for dehydroascorbate reductase directly in potato. Moreover, the authors found that the increment of vitamin C was reached only by overexpressing the cytosolic isoform of the enzyme and not the chloroplastic one, reminding that also the selection of the right protein with the correct motif and signal peptide is fundamental if a specific tissue expression is desired. The work of De Lepeleire et al. (2018) is a smart example of synthetic biology approach in potato. By combining *Arabidopsis* and rice genes encoding for mitochondrial enzymes of folate biosynthetic pathway, under the control of a specific tuber promoter, they managed to get up to 12-fold increase in folate content in mature tubers. A similar result is quite difficult to reach by exploiting the natural variation of potato folate content (De Lepeleire et al. 2018).

The four most significant minerals in terms of nutrient deficiency are Fe, Zn, Si, and Ca. In the last decades, there has been an increasing interest in the development of crops tailored to provide all these limiting nutrients simultaneously. In potato, much attention has been drawn on Ca. In this context, the target genes for Ca increase in tubers are Ca2+ transporters such as calcium exchanger 1 (*CAX1*) and 2 (*CAX2*). Kim et al. (2006) induced an activation of *A. thaliana* CAX2 in *S. tuberosum* var Daejiree obtaining tubers containing 50–65% more Ca than wild-type tubers. Concerning Fe accumulation in potato, nowadays there are no studies on activation or repression of genes for Fe accumulation. Singh et al. (2022) suggested a list of candidate genes for future studies. Very few metabolic engineering researches aimed to biofortify potato tubers with specialized metabolites like anthocyanins and, generally, phenylpropanoids. Indeed, as discussed in previous paragraphs, purple potato varieties already have an important amount of anthocyanins and other phenolics like chlorogenic acids. Moreover, these compounds become quite stable at domestic cooking method (D'Amelia et al. 2022). Most of the studies on purple potatoes were conducted for the characterization of genes encoding for key enzymes of the phenylpropanoid pathway. Among the *plethora* of studies available, Kostyn and collaborators (Kostyn et al. 2013) were the few which tested the nutritional power of DFR-transgenic and non-transgenic *S. tuberosum* cv. Desiree tubers on 20 rats. Their results demonstrated the positive impact of flavonoids on lipid profile. They observed that the level of toxic glycoalkaloids (α-chaconine and α-solanine) increased for about 70% in the DFR-overexpressing tubers. As regards other compounds with nutraceutical value, there are very few studies in potato. For example, lipids (and in particular fatty acids) have been principally studied to increase tolerance to biotic and abiotic stressors. To the best of authors' knowledge, antisense-mediated silencing of lipoxygenase isoforms has been used in *S. tuberosum* to increase tolerance to insects, while no studies for bio-fortification of tubers have been released (Table 4).

5.2 Targeted and Innovative Methods: Organelle Transformation and Genome/Gene Editing

Over the past decades, the expression of transgenes in the plastid genome attracted biotechnologists because it offers several potential advantages compared to the

transformation of the nuclear genome, such as the precise integration of transgenes in the host genome by homologous recombination, the high expression levels and the biological containment of transgenes and recombinant products, the cellular compartmentalization of compounds harmful to the plant, and the possibility of co-expressing several transgenes in prokaryotic like operons. Notwithstanding the potential advantages of this technology, not all attempts have been successful. Many factors (e.g., choice of expression elements, coding region, integration sites, etc.) and parameters (e.g., host plant, growth conditions, tissue, etc.) can affect the achievement of a promising result. In most crops, the efficiency of the tissue culture, selection, and regeneration procedures is considered the most serious bottleneck to plastid transformation (Scotti et al. 2013). Another important limitation in some crops is the generally inefficient gene expression in nongreen plastids, due to deficiencies in the expression machinery of nongreen plastids compared to leaf chloroplasts (Valkov et al. 2009). The first attempts of potato plastid transformation were carried out using vectors developed for tobacco and are characterized by low transformation frequencies and low transgene expression in tubers of transplastomic plants (Thanh et al. 2005). A strong increase in plastid transformation efficiency in potato (one shoot per bombardment), comparable to those obtained with tobacco, was achieved using a combined strategy based on an optimized selection/regeneration procedure and species-specific vectors (Fig. 3) containing potato homologous flanking sequences and regulatory sequences to improve transgene (*gfp*) integration and expression in amyloplasts (Valkov et al. 2021). Although protein accumulation in amyloplasts was low, the GFP protein obtained in tubers using *clpP* regulatory sequences (i.e., promoter and 5′ untranslated region, UTR) confirms its positive effects on transgene translation in

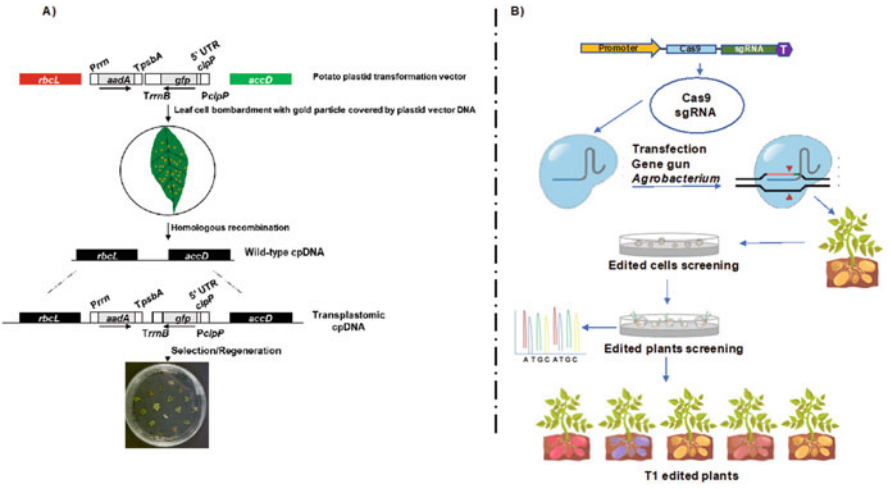


Fig. 3 Schematic representation of unconventional and innovative methods for metabolic engineering. (**a**) Improved chloroplast transformation in potato based on the use of an optimized selection/regeneration procedure and species-specific vector containing potato homologous flanking and regulatory sequences. (**b**) Gene editing strategy based on CRISPR/Cas9 and *Agrobacterium*-mediated transformation

nongreen plastids, with *clpP* being one of the less downregulated genes in tubers compared to leaves (Valkov et al. 2009). To increase protein accumulation (up to 1.3% total soluble protein, TSP) in potato amyloplasts, Yu et al. (2019) developed a different regulatory approach based on a variant of the maize RNA-binding protein PPR10, which activates the expression of the plastid *atpH* gene by binding to a cis-element in the *atpH* 5′ UTR increasing the translation efficiency, and a cognate binding site upstream of *gfp*. Attempts to increase other nutraceutical molecules by means of organelle transformation have been made in particular for specialized metabolites. Indeed, several metabolic pathways occur within the plastids producing, for example, precursors of phenylpropanoid and isoprenoids (through, respectively, the shikimate and methylerythritol 4-phosphate pathways). In tomato, a closely related potato species, Lu et al. (2013) obtained an increase of up to tenfold in vitamin E by including, in the synthetic operon construct, a binding site for an RNA-binding protein from the pentatricopeptide repeat (PPR) family (the intercistronic expression element, IEE). Methods for organelle transformation are not of recent origin, but these can be considered more unconventional than nuclear transgenesis and with a great potential for metabolic engineering of potato. The above reported results suggest that the use of these regulatory sequences may be sufficient to manipulate the expression of enzymatic proteins for metabolic engineering purposes in order to biofortify potato tubers.

More innovative for plant is the genome editing. A group of advanced biotechnological tools are becoming very useful for targeted mutagenesis. These tools can be fully exploited in crop also to improve the production of nutraceuticals through gene knockout or by modifying nucleotide sequences (Hameed et al. 2018). Among these technologies, today, much attention is being paid to the RNA-guided nuclease clustered regularly interspaced short palindromic repeats (CRISPR)-associated (Cas), which is handier technology than to previous methods of editing such as zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs). The CRISPR/Cas system has two main components, the single guide RNA (sgRNA) and the endonuclease Cas. The former is complementary to the target DNA sequence that must be immediately followed by an adjacent protospacer motif recognized by the Cas protein. Once the CRISPR/Cas system recognizes the target DNA, Cas generates a double-stranded break that can be repaired by the non-homologous end-joining (NHEJ) or by the homology-directed recombination (HDR). In particular, NHEJ causes random mutations, while HDR can introduce specific mutations with the help of an additionally provided donor DNA carrying the desired mutation (reviewed by Gonzalez-Salinas et al. 2022). There are few examples of applications of CRISPR/Cas-mediated mutagenesis, and in general of the gene editing approach, addressed to positively impact the nutraceutical potential of potato tubers. Recent studies have been mainly focused to improve the efficiency of these technologies in potato. The gene *GBSSI*, encoding for granule bond starch synthase I enzyme, has been particularly targeted to assess the technology. *GBSSI* is a single copy gene and its knockout produces an easily identifiable phenotype (amylose-free starch). From the nutraceutical point of view, the low value of amylose/amylopectin ratio caused by the de-functionalization of *GBSSI* is not a desirable trait since it is associated with a higher glycemic index of the starch. In this regard, results obtained by the CRISPR/Cas9 editing approach of Tuncel

et al. (2019) can be considered of nutraceutical interest since the authors managed to edit the gene *SBE* (Y08786) that, opposite to *GBBSI*, is responsible for amylopectin formation in tubers and consequently produces an increment of the amylose/amylopectin ratio. Zheng et al. (2021) mutated the sterol side chain reductase 2 encoding gene (*SSR2*) obtaining a reduction of more than half of the content of toxic steroidal glycoalkaloids accumulated in the peel and the tubers of wild-type potato.

Several works have fixed important technical features that will allow to perform efficient CRISPR/Cas-mediated mutagenesis in potato. A doubling in editing frequency of regenerated potato plants has been reached by using endogenous potato U6 promoter, driving the expression of RNA guides, rather than the standard *Arabidopsis* one (Andersson et al. 2017). By using the endogenous promoter to drive CRISPR/Cas components, mutation in all four alleles for 35% tetraploid ex plants was also observed (Johansen et al. 2019). Indeed, the chance to get mutation in all four alleles during the first edited generation is fundamental to reduce or even avoid selfing or backcrossing which is more complicated in tetraploid potato. Furthermore, crossings are also needed to eliminate vector cassette containing both nuclease and the sgRNA. A successful alternative approach consists in the delivering of CRISPR/Cas9 as ribonucleoproteins (RNPs) into cells, with the added potential of generating transgene-free targeted genome edits efficiently (Woo et al. 2015). Indeed, RNPs are more specific than vectors delivering CRISPR/Cas and act more rapidly because they do not need intracellular transcription and translation. A clear reduction in off-target mutations has also widely been demonstrated. Among all, González et al. (2020) developed potato varieties with reduced enzymatic browning in tubers by the specific RNP-based editing of a single member of the *StPPO* gene family (polyphenol oxidases). Their selected edited lines displayed mutations only in the *StPPO2* gene (U22921.1), with no alteration in the coding sequences of other members of the *StPPO* gene family. Another peculiar strength of RNP-based editing is the avoidance of transgene integration in the cellular genome, which reduces the regulatory burden in some countries and allows the edited plants to be labeled as "GMO-free." But "all that is gold does not glitter"! The weaknesses of RNP-based editing are the protoplast transformation methods, which are laborious and require expensive enzymes, and the onset somaclonal variations in edited plants due to much longer culture periods. However, in comparison with the protoplast method, the *Agrobacterium* system (used for CRISPR/Cas vector delivering) is less expensive and needs shorter time frame for regenerant production, reducing the occurrence of somaclonal variations. For this reason, the scientific community is moving forward a CRISPR/Cas variant with reduced DNA cleavage off-targets such as the SuperFi-Cas9 (Bravo et al. 2022).

6 Future Challenges for Increasing Nutraceutical Molecules in Potato

Potato is a perfect candidate for nutrient bio-fortification. The literature reviewed showed that there has been great deal of interest in characterizing the variability of the biochemical composition of potato tubers, but genomic and biotechnological

studies dedicated to biofortify potato tubers are still limited compared to other relative species. For example, relatively few genome-wide studies have been dedicated to identify and characterize structural variants of genes associated to the production of specific metabolites, and most of these concern with the anthocyanin biosynthesis. Association between genomic variants and amount of minerals, vitamins, or carotenoids is still rare, limiting the development of specific markers to improve the content of these compounds by classical breeding approaches. Biotechnological reports (i.e., metabolic engineering approaches) are also relatively few. In our opinion, there are three main challenges that should be considered when developing new breeding programs addressed to potato bio-fortification:

- A considerable increment of nutraceuticals is needed for potato, compared to fresh consumed vegetables, to counterbalance the reduction caused by heat treatments. Potato tubers need to be cooked before consumption, and cooking methods may degrade phytochemicals with nutraceutical properties. There are different studies that checked the stability of these phytochemicals in tuber skin and flesh after different domestic cooking methods (D'Amelia et al. 2022). Anthocyanins and chlorogenic acids have been reported to be stable, while other compounds, such as vitamins (vitamin C and folate) and flavonols, showed a consistent reduction in the total amount (Stewart and Taylor 2017). The storage of tubers may also lead to the reduction of instable compounds. Therefore, another challenge in the field of tuber bio-fortification regards the possibility to drive post-biosynthetic modification of main molecules to stabilize them over time. For example, it is known that hydroxylation negatively influences anthocyanin stability; instead, methoxylation/acylation positively affects stability.
- There are several side effects evaluated in the bio-fortification of the tuber flesh. For example, a high tendency of tuber flesh with improved CGA to browning during cutting and cooking because of CGA oxidation has been observed (Ali et al. 2016). It has also been observed that the level of toxic glycoalkaloids (α-chaconine and α-solanine) increased by about 70%, along with anthocyanins, in the DFR-overexpressing tubers (Kostyn et al. 2013). Hence, an accurate selection of the promoter, besides genes and relative alleles, is important to drive a well-balanced expression.
- The tetraploid genetic asset of the cultivated potato *S. tuberosum* is an aspect which discourages breeders and biotechnologists working in the field of food bio-fortification. The tetrasomic inheritance makes genetic studies and breeding programs addressed to introduce new traits quite complicated and time consuming compared to those of a diploid plant. There are also additional noteworthy constraints. In fact, depending on gene-centromere distance, alternative segregation patterns in the gametes are possible. Besides normal chromosome segregation, the possibility exists that portions of two sister chromatids merge in the same gamete. This occurs when recombination happens between a locus and the centromere. Due to this phenomenon (called double reduction), a triplex AAAa individual can produce aa gametes, with a frequency proportional to the distance between the locus and the centromere. In addition, as any clonal crop, the

cultivated potato is highly heterozygous, and homozygous genotypes are difficult to obtain upon selfing due to inbreeding depression. The recent development of self-compatible diploid potatoes makes it possible to obtain homozygous parental lines to use in creating F1 hybrid diploid potatoes (Zhang et al. 2021).

References

Acharjee A, Chibon PY, Kloosterman B, America T, Renaut J, Maliepaard C, Visser RG (2018) Genetical genomics of quality related traits in potato tubers using proteomics. BMC Plant Biology 18(1):1–10

Ahn JY, Kim J, Yang JY, Lee HJ, Kim S, Cho KS, Lee SH, Kim JH, Lee TH, Hur Y, Shim D (2022) Comparative transcriptome analysis between two potato cultivars in tuber induction to reveal associated genes with anthocyanin accumulation. Int J Mol Sci 23:3681. https://doi.org/10.3390/ijms23073681

Ali HM, El-Gizawy AM, El-Bassiouny REI, Saleh M (2016) The role of various amino acids in enzymatic browning process in potato tubers, and identifying the browning products. Food Chem 192:879–885. https://doi.org/10.1016/j.foodchem.2015.07.100

Andersson M, Turesson H, Nicolia A, Fält AS, Samuelsson M, Hofvander P (2017) Efficient targeted multiallelic mutagenesis in tetraploid potato (*Solanum tuberosum*) by transient CRISPR-Cas9 expression in protoplasts. Plant Cell Rep 36:117–128. https://doi.org/10.1007/s00299-016-2062-3

Azeem F, Hussain M, Hussain S, Zubair M, Nadeem H, Ali MA, Afzal M, Siddique MH (2021) Genome-wide analysis and expression profiling of potassium transport related genes in *Solanum tuberosum*. Pakistan J Agric Sci 58:1. https://doi.org/10.21162/PAKJAS/21.9703

Bagri DS, Upadhyaya DC, Kumar A, Upadhyaya CP (2018) Overexpression of PDX-II gene in potato (*Solanum tuberosum* L.) leads to the enhanced accumulation of vitamin B6 in tuber tissues and tolerance to abiotic stresses. Plant Sci 272:267–275. https://doi.org/10.1016/j.plantsci.2018.04.024

Bali S, Robinson BR, Sathuvalli V, Bamberg J, Goyer A (2018) Single nucleotide polymorphism (SNP) markers associated with high folate content in wild potato species. PLoS One 13:e0193415. https://doi.org/10.1371/journal.pone.0193415

Bao Y, Nie T, Wang D, Chen Q (2022) Anthocyanin regulatory networks in *Solanum tuberosum* L. leaves elucidated via integrated metabolomics, transcriptomics, and StAN1 overexpression. BMC Plant Biol 22:1–17. https://doi.org/10.1186/s12870-022-03557-1

Barlagne C, Kyle C, Kuhfuss L, Johan-lagerkvist C (2021) Consumers' preferences and perceptions of genetic diversity and health benefits of potatoes summary: RESAS WP232. Prot Genet Divers Deliv O36c:1–13

Bártová V, Bárta J, Jarošová M (2019) Antifungal and antimicrobial proteins and peptides of potato (*Solanum tuberosum* L.) tubers and their applications. Appl Microbiol Biotechnol 103:5533–5547. https://doi.org/10.1007/s00253-019-09887-9

Blancquaert D, Storozhenko S, Van Daele J, Stove C, Visser RG, Lambert W, Van Der Straeten D (2013) Enhancing pterin and para-aminobenzoate content is not sufficient to successfully biofortify potato tubers and *Arabidopsis thaliana* plants with folate. J Exp Bot 64:3899–3909. https://doi.org/10.1093/jxb/ert224

Bonar N, Liney M, Zhang R, Austin C, Dessoly J, Davidson D, Stephens J, McDougall G, Taylor M, Bryan GJ, Hornyik C (2018) Potato miR828 is associated with purple tuber skin and flesh color. Front Plant Sci 871:1742. https://doi.org/10.3389/fpls.2018.01742

Bonierbale MW, Amoros WR, Salas E, De Jong W (2020) Potato breeding. In: Campos H, Ortiz O (eds) The potato crop: its agricultural, nutritional and social contribution to humankind. Springer International Publishing. Springer, Cham, pp 163–217

Boudaba N, Marion A, Huet C, Pierre R, Viollet B, Foretz M (2018) AMPK re-activation suppresses hepatic steatosis but its downregulation does not promote fatty liver development. EBioMedicine 28:194–209. https://doi.org/10.1016/j.ebiom.2018.01.008

Bradshaw JE (2019) Improving the nutritional value of potatoes by conventional breeding and genetic modification. In: Quality breeding in field crops. Springer International Publishing, Cham, pp 41–84

Bravo JP, Liu MS, Hibshman GN, Dangerfield TL, Jung K, McCool RS, Johnson KA, Taylor DW (2022) Structural basis for mismatch surveillance by CRISPR-Cas9. Nature 603:343–347. https://doi.org/10.1038/s41586-022-04470-1

Brown CR, Edwards CG, Yang C-P, Dean BB (2019) Orange flesh trait in potato: inheritance and carotenoid content. J Am Soc Hortic Sci 118:145–150. https://doi.org/10.21273/jashs.118.1.145

Burgos G, Zum Felde T, Andre C, Kubow S (2020) The potato and its contribution to the human diet and health. In: The potato crop. Springer International Publishing, Cham, pp 37–74. https://doi.org/10.1007/978-3-030-28683-5_2

Campbell R, Pont SDA, Morris JA, McKenzie G, Sharma SK, Hedley PE, Ramsay G, Bryan GJ, Taylor MA (2014) Genome-wide QTL and bulked transcriptomic analysis reveals new candidate genes for the control of tuber carotenoid content in potato (*Solanum tuberosum* L.). Theor Appl Genet 127(9):1917–1933. https://doi.org/10.1007/s00122-014-2349-0

Campbell R, Morris WL, Mortimer CL, Misawa N, Ducreux LJ, Morris JA, Hedley PE, Fraser PD, Taylor MA (2015) Optimising ketocarotenoid production in potato tubers: effect of genetic background, transgene combinations and environment. Plant Sci 234:27–37. https://doi.org/10.1016/j.plantsci.2015.01.014

Cao M, Liu H, Zhang C, Wang D, Liu X, Chen Q (2020) Functional analysis of StPHT1;7, a *Solanum tuberosum* L. phosphate transporter gene, in growth and drought tolerance. Plan Theory 9:1384. https://doi.org/10.3390/plants9101384

Cárdenas PD, Sonawane PD, Heinig U, Jozwiak A, Panda S, Abebie B, Kazachkova Y, Pliner M, Unger T, Wolf D, Ofner I (2019) Pathways to defense metabolites and evading fruit bitterness in genus Solanum evolved through 2-oxoglutarate-dependent dioxygenases. Nat Commun 10:1–13. https://doi.org/10.1038/s41467-019-13211-4

Crowell EF, McGrath JM, Douches DS (2008) Accumulation of vitamin E in potato (*Solanum tuberosum*) tubers. Transgenic Res 17:205–217. https://doi.org/10.1007/s11248-007-9091-1

D'Amelia V, Aversano R, Ruggiero A, Batelli G, Appelhagen I, Dinacci C, Hill L, Martin C, Carputo D (2018) Subfunctionalization of duplicate MYB genes in *Solanum commersonii* generated the cold-induced ScAN2 and the anthocyanin regulator ScAN1. Plant Cell Environ 41:1038–1051. https://doi.org/10.1111/pce.12966

D'Amelia V, Raiola A, Carputo D, Filippone E, Barone A, Rigano MM (2019) A basic Helix-Loop-Helix (SlARANCIO), identified from a *Solanum pennellii* introgression line, affects carotenoid accumulation in tomato fruits. Sci Rep 9:1–10. https://doi.org/10.1038/s41598-019-40142-3

D'Amelia V, Sarais G, Fais G, Dessì D, Giannini V, Garramone R, Carputo D, Melito S (2022) Biochemical characterization and effects of cooking methods on main phytochemicals of red and purple potato tubers, a natural functional food. Foods 11:384. https://doi.org/10.3390/foods11030384

De Jong WS, Eannetta NT, De Jong DM, Bodis M (2004) Candidate gene analysis of anthocyanin pigmentation loci in the Solanaceae. Theor Appl Genet 108:423–432. https://doi.org/10.1007/s00122-003-1455-1

De Lepeleire J, Strobbe S, Verstraete J, Blancquaert D, Ambach L, Visser RG, Stove C, Van Der Straeten D (2018) Folate biofortification of potato by tuber-specific expression of four folate biosynthesis genes. Mol Plant 11:175–188. https://doi.org/10.1016/j.molp.2017.12.008

Diretto G, Tavazza R, Welsch R, Pizzichini D, Mourgues F, Papacchioli V, Beyer P, Giuliano G (2006) Metabolic engineering of potato tuber carotenoids through tuber-specific silencing of lycopene epsilon cyclase. BMC Plant Biol 6:1–11. https://doi.org/10.1186/1471-2229-6-13

Diretto G, Al-Babili S, Tavazza R, Papacchioli V, Beyer P, Giuliano G (2007a) Metabolic engineering of potato carotenoid content through tuber-specific overexpression of a bacterial mini-pathway. PLoS One 2:e350. https://doi.org/10.1371/journal.pone.0000350

Diretto G, Welsch R, Tavazza R, Mourgues F, Pizzichini D, Beyer P, Giuliano G (2007b) Silencing of beta-carotene hydroxylase increases total carotenoid and beta-carotene levels in potato tubers. BMC Plant Biol 7:1–8. https://doi.org/10.1186/1471-2229-7-11

Dolničar P (2021) Importance of potato as a crop and practical approaches to potato breeding. In: Methods in molecular biology. Humana, New York, pp 3–20

Duarte-Delgado D, Ñústez-López CE, Narváez-Cuenca CE et al (2016) Natural variation of sucrose, glucose and fructose contents in Colombian genotypes of *Solanum tuberosum* Group *Phureja* at harvest. J Sci Food Agric 96:4288–4294. https://doi.org/10.1002/jsfa.7783

Ducreux LJM, Morris WL, Hedley PE et al (2005) Metabolic engineering of high carotenoid potato tubers containing enhanced levels of β-carotene and lutein. J Exp Bot 56:81–89. https://doi.org/10.1093/jxb/eri016

EFSA NDA Panel (EFSA Panel on Dietetic Products N and A) (2015) Scientific opinion on dietary reference values for calcium. EFSA J 13:1–82. https://doi.org/10.2903/j.efsa.2015.4101

Eschen-Lippold L, Rothe G, Stumpe M et al (2007) Reduction of divinyl ether-containing polyunsaturated fatty acids in transgenic potato plants. Phytochemistry 68:797–801. https://doi.org/10.1016/j.phytochem.2006.12.010

Fu LS, Doreswamy V, Prakash R (2014) The biochemical pathways of central nervous system neural degeneration in niacin deficiency. Neural Regen Res 9:1509. https://doi.org/10.4103/1673-5374.139475

Göbel C, Feussner I, Rosahl S (2003) Lipid peroxidation during the hypersensitive response in potato in the absence of 9-lipoxygenases. J Biol Chem 278:52834–52840. https://doi.org/10.1074/jbc.M310833200

González MN, Massa GA, Andersson M et al (2020) Reduced enzymatic browning in potato tubers by specific editing of a polyphenol oxidase gene via ribonucleoprotein complexes delivery of the Crispr/Cas9 system. Front Plant Sci 10:1649. https://doi.org/10.3389/fpls.2019.01649

Gonzalez-Salinas F, Martinez-Amador C, Trevino V (2022) Characterizing genes associated with cancer using the CRISPR/Cas9 system: a systematic review of genes and methodological approaches. Gene 833:146595. https://doi.org/10.1016/j.gene.2022.146595

Goo YM, Hyun JC, Kim TW et al (2008) Expressional characterization of dehydroascorbate reductase cDNA in transgenic potato plants. J Plant Biol 51:35–41. https://doi.org/10.1007/BF03030738

Hackett CA, Boskamp B, Vogogias A, Preedy KF, Milne I (2017) TetraploidSNPMap: software for linkage analysis and QTL mapping in autotetraploid populations using SNP dosage data. J Hered 108:438–442. https://doi.org/10.1093/jhered/esx022

Hameed A, Zaidi SS-A, Shakir S, Mansoor S (2018) Applications of new breeding technologies for potato improvement. Front Plant Sci 9:853–871. https://doi.org/10.3389/fpls.2018.00925

He G, Qin L, Tian W, Meng L, He T, Zhao D (2020) Heavy metal transporters-associated proteins in *S. tuberosum*: genome-wide identification, comprehensive gene feature, evolution and expression analysis. Genes 11:1–21. https://doi.org/10.3390/genes11111269

Hellmann H, Goyer A, Navarre DA (2021) Antioxidants in potatoes: a functional view on one of the major food crops worldwide. Molecules 26:2446. https://doi.org/10.3390/molecules26092446

Hussain M, Qayum A, Xiuxiu Z, Liu L, Hussain K, Yue P, Yue S, Koko MY, Hussain A, Li X (2021) Potato protein: an emerging source of high quality and allergy free protein, and its possible future based products. Food Res Int 148:110583. https://doi.org/10.1016/j.foodres.2021.110583

Ierna A, Pellegrino A, Mauro RP, Leonardi C (2020) Micronutrient foliar fertilization for the biofortification of raw and minimally processed early potatoes. Agronomy 10:1744. https://doi.org/10.3390/agronomy10111744

Johansen IE, Liu Y, Jørgensen B, Bennett EP, Andreasson E, Nielsen KL, Blennow A, Petersen BL (2019) High efficacy full allelic CRISPR/Cas9 gene editing in tetraploid potato. Sci Rep 9:1–7. https://doi.org/10.1038/s41598-019-54126-w

Kärenlampi SO, White PJ (2009) Potato proteins, lipids, and minerals. In: Advances in potato chemistry and technology. Elsevier, New York, pp 99–125. https://doi.org/10.1016/B978-0-12-374349-7.00005-2

Kawar PG (2016) Enriched potato for mitigating hidden hunger. In: Biofortification of food crops. Springer, New Delhi, pp 433–457

Kim CK, Han JS, Lee HS, Oh JY, Shigaki T, Park SH, Hirschi K (2006) Expression of an Arabidopsis CAX2 variant in potato tubers increases calcium levels with no accumulation of manganese. Plant Cell Rep 25:1226–1232. https://doi.org/10.1007/s00299-006-0214-6

Klaassen MT, Bourke PM, Maliepaard C, Trindade LM (2019a) Multi-allelic QTL analysis of protein content in a bi-parental population of cultivated tetraploid potato. Euphytica 215:1–18. https://doi.org/10.1007/s10681-018-2331-z

Klaassen MT, Willemsen JH, Vos PG, Visser RG, van Eck HJ, Maliepaard C, Trindade LM (2019b) Genome-wide association analysis in tetraploid potato reveals four QTLs for protein content. Mol Breed 39:1–12. https://doi.org/10.1007/s11032-019-1070-8

Komarnytsky S, Cook A, Raskin I (2011) Potato protease inhibitors inhibit food intake and increase circulating cholecystokinin levels by a trypsin-dependent mechanism. Int J Obes 35:236–243. https://doi.org/10.1038/ijo.2010.192

Kostyn K, Szatkowski M, Kulma A, Kosieradzka I, Szopa J (2013) Transgenic potato plants with overexpression of dihydroflavonol reductase can serve as efficient nutrition sources. J Agric Food Chem 61:6743–6753. https://doi.org/10.1021/jf400645s

Kowalczewski PŁ, Olejnik A, Białas W, Rybicka I, Zielińska-Dawidziak M, Siger A, Kubiak P, Lewandowicz G (2019) The nutritional value and biological activity of concentrated protein fraction of potato juice. Nutrients 11:1523. https://doi.org/10.3390/nu11071523

Kromann P, Valverde F, Alvarado S, Vélez R, Pisuña J, Potosí B, Taipe A, Caballero D, Cabezas A, Devaux A (2017) Can Andean potatoes be agronomically biofortified with iron and zinc fertilizers? Plant Soil 411:121–138. https://doi.org/10.1007/s11104-016-3065-0

Lehesranta SJ, Davies HV, Shepherd LV, Nunan N, McNicol JW, Auriola S, Koistinen KM, Suomalainen S, Kokko HI, Karenlampi SO (2005) Comparison of tuber proteomes of potato varieties, landraces, and genetically modified lines. Plant Physiol 138:1690–1699. https://doi.org/10.1104/pp.105.060152

Li XB, Suo HC, Liu JT, Wang L, Li CC, Liu W (2020) Genome-wide identification and expression analysis of the potato ZIP gene family under Zn-deficiency. Biol Plant 64:845–855. https://doi.org/10.32615/bp.2020.125

Liu Q, Guo Q, Akbar S, Zhi Y, El Tahchy A, Mitchell M, Li Z, Shrestha P, Vanhercke T, Ral JP, Liang G (2016) Genetic enhancement of oil content in potato tuber (*Solanum tuberosum* L.) through an integrated metabolic engineering strategy. Plant Biotechnol J 5:56–67. https://doi.org/10.1111/pbi.12590

Liyanage R, Minamino S, Nakamura Y, Shimada KI, Sekikawa M, Sasaki K, Ohba K, Jayawardana BC, Shibayama SI, Fukushima M (2010) Preparation method modulates hypocholesterolaemic responses of potato peptides. J Funct Foods 2:118–125. https://doi.org/10.1016/j.jff.2010.03.001

Liyanage DW, Yevtushenko DP, Konschuh M, Bizimungu B, Lu ZX (2021) Processing strategies to decrease acrylamide formation, reducing sugars and free asparagine content in potato chips from three commercial cultivars. Food Control 119:107452. https://doi.org/10.1016/j.foodcont.2020.107452

Lopez AB, Van Eck J, Conlin BJ, Paolillo DJ, O'Neill J, Li L (2008) Effect of the cauliflower or transgene on carotenoid accumulation and chromoplast formation in transgenic potato tubers. J Exp Bot 59:213–223. https://doi.org/10.1093/jxb/erm299

Lu S, Van Eck J, Zhou X, Lopez AB, O'Halloran DM, Cosman KM, Conlin BJ, Paolillo DJ, Garvin DF, Vrebalov J, Kochian LV (2007) The cauliflower or gene encodes a dnaj cysteine-rich domain-containing protein that mediates high levels of β-carotene accumulation. Plant Cell 18:3594–3605. https://doi.org/10.1105/tpc.106.046417

Lu Y, Rijzaani H, Karcher D, Ruf S, Bock R (2013) Efficient metabolic pathway engineering in transgenic tobacco and tomato plastids with synthetic multigene operons. Proc Natl Acad Sci U S A 110:623–632. https://doi.org/10.1073/pnas.1216898110

Lu H, Tian Z, Cui Y, Liu Z, Ma X (2020) Chlorogenic acid: a comprehensive review of the dietary sources, processing effects, bioavailability, beneficial properties, mechanisms of action, and future directions. Compr Rev Food Sci Food Saf 19:3130–3158. https://doi.org/10.1111/1541-4337.12620

Meenu Krishnan VG, Murugan K (2016) Solanum protease inhibitors and their therapeutic potentialities: a review. Int J Pharm Pharm Sci 8:14–21. https://doi.org/10.22159/ijpps.2016v8i12.14836

Mengist MF, Alves S, Griffin D, Creedon J, McLaughlin MJ, Jones PW, Milbourne D (2018) Genetic mapping of quantitative trait loci for tuber-cadmium and zinc concentration in potato reveals associations with maturity and both overlapping and independent components of genetic control. Theor Appl Genet 131:929–945. https://doi.org/10.1007/s00122-017-3048-4

Mi X, Ji X, Yang J, Liang L, Si H, Wu J, Zhang N, Wang D (2015) Transgenic potato plants expressing *cry3A* gene confer resistance to Colorado potato beetle. Comptes Rendus - Biol 338: 443–450. https://doi.org/10.1016/j.crvi.2015.04.005

Milczarek D, Tatarowska B (2018) Evaluation of potato cultivars and breeding lines for carotenoids content in tubers. Plant Breed Seed Sci 75:11–16. https://doi.org/10.1515/plass-2017-0003

Mooney S, Chen L, Kühn C, Navarre R, Knowles NR, Hellmann H (2013) Genotype-specific changes in vitamin B6 content and the PDX family in potato. Biomed Res Int 2013:389723. https://doi.org/10.1155/2013/389723

Naeem M, Demirel U, Yousaf MF, Caliskan S, Caliskan ME (2021) Overview on domestication, breeding, genetic gain and improvement of tuber quality traits of potato using fast forwarding technique (GWAS): a review. Plant Breed 140:519–542

Osorio CE (2019) The role of *Orange* gene in carotenoid accumulation: manipulating chromoplasts toward a colored future. Front Plant Sci 10:1235. https://doi.org/10.3389/fpls.2019.01235

Park S, Kang TS, Kim CK, Han JS, Kim S, Smith RH, Pike LM, Hirschi KD (2005) Genetic manipulation for enhancing calcium content in potato tuber. J Agric Food Chem 53:5598–5603. https://doi.org/10.1021/jf050531c

Qin A, Shi Q, Yu X (2011) Ascorbic acid contents in transgenic potato plants overexpressing two dehydroascorbate reductase genes. Mol Biol Rep 38:1557–1566. https://doi.org/10.1007/s11033-010-0264-2

Rommens CM, Richael CM, Yan H, Navarre DA, Ye J, Krucker M, Swords K (2008) Engineered native pathways for high kaempferol and caffeoylquinate production in potato. Plant Biotechnol J 6:870–886. https://doi.org/10.1111/j.1467-7652.2008.00362.x

Schrenk D, Bignami M, Bodin L, Chipman JK, del Mazo J, Hogstrand C, Hoogenboom L, Leblanc JC, Nebbia CS, Nielsen E (2020) Risk assessment of glycoalkaloids in feed and food, in particular in potatoes and potato-derived products. EFSA J 18:06222. https://doi.org/10.2903/j.efsa.2020.6222

Scotti N, Bellucci M, Cardi T (2013) The chloroplasts as platform for recombinant proteins production. In: Translation in mitochondria and other organelles. Springer, Berlin, Heidelberg, pp 225–262. https://doi.org/10.1007/978-3-642-39426-3_10

Singh B, Raigond P, Dutt S, Kumar M (2020) Potatoes for food and nutritional security. In: Potato. Springer, Singapore, pp 1–12

Singh B, Goutam U, Kukreja S, Siddappa S, Sood S, Sharma J, Bhardwaj V (2022) Biofortification strategies to improve iron concentrations in potato tubers: lessons and future opportunities. Potato Res 65:51–64. https://doi.org/10.1007/s11540-021-09508-x

Sitjà-Arnau M, Molina MA, Blanco-Aparicio C, Ferrer-Soler L, Lorenzo J, Avilés FX, Querol E, De Llorens R (2005) Mechanism of action of potato carboxypeptidase inhibitor (PCI) as an EGF blocker. Cancer Lett 226:169–184. https://doi.org/10.1016/j.canlet.2005.01.025

Sonar CR, Parhi A, Liu F, Patel J, Rasco B, Tang J, Sablani SS (2020) Investigating thermal and storage stability of vitamins in pasteurized mashed potatoes packed in barrier packaging films. Food Packag Shelf Life 24:100486. https://doi.org/10.1016/j.fpsl.2020.100486

Stewart D, Taylor M (2017) Potato–A basis for human nutrition and health benefits. AHDB Potatoes, pp 1–60

Sulli M, Mandolino G, Sturaro M, Onofri C, Diretto G, Parisi B, Giuliano G (2017) Molecular and biochemical characterization of a potato collection with contrasting tuber carotenoid content. PLoS One 12:0184143. https://doi.org/10.1371/journal.pone.0184143

Tang D, Jia Y, Zhang J, Li H, Cheng L, Wang P, Bao Z, Liu Z, Feng S, Zhu X, Li D (2022) Genome evolution and diversity of wild and cultivated potatoes. Nature 606:1–7. https://doi.org/10.1038/s41586-022-04822-x

Thanh TN, Nugent G, Cardi T, Dix PJ (2005) Generation of homoplasmic plastid transformants of a commercial cultivar of potato (*Solanum tuberosum* L.). Plant Sci 168:1495–1500. https://doi.org/10.1016/j.plantsci.2005.01.023

Tuncel A, Corbin KR, Ahn-Jarvis J, Harris S, Hawkins E, Smedley MA, Harwood W, Warren FJ, Patron NJ, Smith AM (2019) Cas9-mediated mutagenesis of potato starch-branching enzymes generates a range of tuber starch phenotypes. Plant Biotechnol J 17:2259–2271. https://doi.org/10.1111/pbi.13137

Upadhyaya CP, Bagri DS (2021) Biotechnological approaches for nutritional improvement in potato (*Solanum tuberosum* L.). In: Genome engineering for crop improvement. Wiley, Hoboken, pp 253–280. https://doi.org/10.1002/9781119672425.ch15

Upadhyaya CP, Akula N, Young KE, Chun SC, Kim DH, Park SW (2010) Enhanced ascorbic acid accumulation in transgenic potato confers tolerance to various abiotic stresses. Biotechnol Lett 32:321–330. https://doi.org/10.1007/s10529-009-0140-0

Valkov VT, Scotti N, Kahlau S, MacLean D, Grillo S, Gray JC, Bock R, Cardi T (2009) Genome-wide analysis of plastid gene expression in potato leaf chloroplasts and tuber amyloplasts: transcriptional and posttranscriptional control. Plant Physiol 150:2030–2044. https://doi.org/10.1104/pp.109.140483

Valkov VT, Gargano D, Cardi T, Scotti N (2021) Plastid transformation in potato: an important source of nutrition and industrial materials. In: Chloroplast biotechnology. Humana, New York, pp 247–256. https://doi.org/10.1007/978-1-0716-1472-3_13

Van Der Straeten D, Bhullar NK, De Steur H et al (2020) Multiplying the efficiency and impact of biofortification through metabolic engineering. Nat Commun 11:1–10. https://doi.org/10.1038/s41467-020-19020-4

Vatansever R, Koc I, Ozyigit II, Sen U, Uras ME, Anjum NA, Pereira E, Filiz E (2016) Genome-wide identification and expression analysis of sulfate transporter (SULTR) genes in potato (*Solanum tuberosum* L.). Planta 244:167–1183. https://doi.org/10.1007/s00425-016-2575-6

Venkatesh J, Yu JW, Park SW (2013) Genome-wide analysis and expression profiling of the *Solanum tuberosum* aquaporins. Plant Physiol Biochem 73:392–404. https://doi.org/10.1016/j.plaphy.2013.10.025

Wierzejska R, Wojda B (2020) Folic acid supplementation in pregnancy and prevention of fetal neural tube defects. Przegl Epidemiol 74:362–369. https://doi.org/10.32394/pe.74.29

Woo JW, Kim J, Kwon SI, Corvalán C, Cho SW, Kim H, Kim SG, Kim ST, Choe S, Kim JS (2015) DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nat Biotechnol 33:1162–1164. https://doi.org/10.1038/nbt.3389

Wu J, Wu Q, Yang D et al (2021) Patatin primary structural properties and effects on lipid metabolism. Food Chem 344:128661. https://doi.org/10.1016/j.foodchem.2020.128661

Yang H, Liao Q, Ma L, Luo W, Xiong X, Luo Y, Yang X, Du C, He Y, Li X, Gao D (2021) Features and genetic basis of chlorogenic acid formation in diploid potatoes. Food Chem Mol Sci 3:100039. https://doi.org/10.1016/j.fochms.2021.100039

Yu Q, Barkan A, Maliga P (2019) Engineered RNA-binding protein for transgene activation in non-green plastids. Nature Plants 5:486–490

Zhang C, Yang Z, Tang D, Zhu Y, Wang P, Li D, Zhu G, Xiong X, Shang Y, Li C, Huang S (2021) Genome design of hybrid potato. Cell 184:3873–3883. https://doi.org/10.1016/j.cell.2021.06.006

Zhang F, Qu L, Gu Y, Xu ZH, Xue HW (2022) Resequencing and genome-wide association studies of autotetraploid potato. Mol Hortic 2:6. https://doi.org/10.1186/s43897-022-00027-y

Zheng Z, Ye G, Zhou Y, Pu X, Su W, Wang J (2021) Editing sterol side chain reductase 2 gene (StSSR2) via CRISPR/Cas9 reduces the total steroidal glycoalkaloids in potato. All Life 14:401–413. https://doi.org/10.1080/26895293.2021.1925358

Tomato: Genetics, Genomics, and Breeding of Health-Related Traits

Ibrahim Çelik, Nergiz Gürbüz Çolak, Sami Doğanlar, and Anne Frary

Contents

I. Çelik
Department of Agricultural and Livestock Production, Çal Vocational School of Higher Education, Pamukkale University, Denizli, Turkey
e-mail: icelik@pau.edu.tr

N. Gürbüz Çolak
Plant Science and Technology Application and Research Center, Izmir Institute of Technology, Urla, Turkey
e-mail: nergizgurbuz@iyte.edu.tr

S. Doğanlar (✉)
Plant Science and Technology Application and Research Center, Izmir Institute of Technology, Urla, Turkey

Department of Molecular Biology and Genetics, Izmir Institute of Technology, Urla, Turkey
e-mail: samidoganlar@iyte.edu.tr

A. Frary
Department of Molecular Biology and Genetics, Izmir Institute of Technology, Urla, Turkey
e-mail: annefrary@iyte.edu.tr

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_49

Abstract

Tomato (*Solanum lycopersicum*) is a popular crop due to its versatility and nutritional quality. In addition to its nutritional content, tomato is rich in various phytochemicals that are known to have beneficial effects on human health. These bioactive components include pigments like lycopene and β-carotene, ascorbic acid (vitamin C), phenolic compounds, polyamines, and glycoalkaloids. Tomato metabolites have various bioactivities such as antihypertensive, antioxidant, anticancer, anti-inflammatory, antidiabetic, antiallergenic, antiatherogenic, anti-thrombotic, and antimicrobial effects. Research aimed at improving tomato for many of these specific activities is still in its infancy; however, a foundation of knowledge has been established for health-related (HR) traits in the crop. In this chapter, previous works surveying tomato germplasm for HR traits, conventional breeding, and genetic investigation of these characteristics are described. We also

discuss research aimed at HR gene mapping and isolation as well as efforts to improve these traits via genetic engineering and genome editing.

Keywords

Antioxidants · Ascorbic acid · β-carotene · Glycoalkaloids · Lycopene · Phytochemicals · *Solanum lycopersicum*

1 Introduction

1.1 Agricultural Importance

Tomato (*Solanum lycopersicum*) is in the Solanaceae family which includes potato, eggplant, and pepper. Cultivated tomato evolved from an ancestor with small green tomatoes and its domestication began in Mexico (Peralta and Spooner 2007). Today, the crop is cultivated worldwide because of its versatility, nutritional quality, and economic importance (Kim et al. 2021). Tomato is grown in the field and greenhouse with different cultivars grown for processing and fresh consumption (Campos et al. 2021). Cultivars vary widely for shape, size, and color (Fig. 1). According to Food and Agriculture Organization (FAO) statistics, the annual global production of tomatoes was approximately 187 million tons in 2020 (https://www.fao.org/faostat/en/#data). Tomato is an important part of the diet and is one of the most consumed vegetable crops worldwide. It is used fresh or processed as sauce, soup, juice, or paste (Ali et al. 2020).

Fig. 1 Example of the wide array of phenotypic diversity available in cultivated tomato. Color, size, and shape variation have been associated with some HR traits. (Photo credit: Mehmet Ülger, MULTI Tarim Seed Co.)

For many years, the improvement of agronomic traits, especially yield, was the main focus of tomato breeding. For this reason, breeding programs often concentrated on development of disease-resistant cultivars (Kim et al. 2021). However, yield gains are often associated with decreased fruit quality (Goff and Klee 2006). More recently, quality traits such as color, flavor, and nutritional quality have become popular breeding targets due to consumer demand and the knowledge that fruits and vegetables make significant contributions to human health (Kusano and Fukushima 2013). Although consumer demands are constantly changing, most people want to consume more nutritious products that also contribute to better health and well-being. This increased interest in health-related (HR) traits in tomato has galvanized the scientific community to gain a greater understanding of the underlying genetic factors and metabolic pathways that control HR traits.

Tomato is a model species for scientific research due to its moderate genome size (950 Mb), diploidy, self-fertility, ease of manipulation, growth under different cultivation conditions, and relatively short life cycle (Kim et al. 2014; Campos et al. 2021). Moreover, the tomato genome has been sequenced and re-sequenced and these data are available online (https://solgenomics.net/).

1.2 Nutritional Composition

A balanced diet consists of certain amounts of vitamins, minerals, proteins, and fats, many of which are found in plants (Ali et al. 2020). In addition, plants are rich in bioactive phytochemicals which have been used to treat or prevent several human diseases (Rao 2003; Giampieri and Battino 2020). Tomato is a good source for the human diet because it is rich in both nutritional and bioactive compounds (Table 1). The nutritional content of tomato mainly consists of a critical amount of soluble and insoluble dietary fibers, carotenoids, phenolic acids, vitamins, proteins, minerals, essential and other amino acids, and fatty acids including monounsaturated fatty acids (Delzenne et al. 2020; Merenkova et al. 2020; Ali et al. 2020). In addition to these nutritional compounds, tomato organs including the fruit contain alkaloids which are antinutritional bioactive compounds (Chaudhary et al. 2018). Tomatoes are considered part of a healthy diet because they contain low fat and cholesterol and harbor significant amounts of specific HR compounds such as lycopene, β-carotene, certain phenolic acids, vitamins C and A, folate, and potassium (Tan et al. 2010). The presence of these HR compounds makes tomato a good source of protection against some diseases such as blindness, respiratory disorders, cardiovascular diseases, and some forms of cancers (Chaudhary et al. 2018). These phytochemicals can also help prevent mutations in DNA and have anti-inflammatory, antimicrobial, and other beneficial activities.

The content of tomato fruit is not constant and varies with many factors like variety, growth conditions, developmental and ripening stage, environmental conditions, and post-harvest and storage conditions (Hall et al. 2008; Manganaris et al. 2018; Boz and Sand 2020). For instance, water-soluble antioxidant activity is high when the fruit are green and decreases when the fruit are ripe, while lipid-soluble antioxidant activity, which is mostly related to lycopene content, increases at the red ripe stage (Cano et al. 2003). Lutein is found at high levels at the red stage but not when fruit are unripe

Table 1 Nutritional composition of 100 g raw, red, ripe tomatoes (USDA FoodData Central)

Compound	Amount
Proximates	
Water	94.5 g
Energy	18 kcal, 74 kJ
Protein	0.88 g
Total lipids	0.2 g
Ash	0.5 g
Carbohydrates	3.89 g
Dietary fiber, total	1.2 g
Sugars, total	2.63 g
Minerals	
Calcium	10 mg
Iron	0.27 mg
Magnesium	11 mg
Phosphorus	24 mg
Potassium	237 mg
Sodium	5 mg
Zinc	0.17 mg
Copper	0.059 mg
Manganese	0.114 mg
Fluoride	2.3 μg
Vitamins	
Ascorbic acid, total	13.7 mg
Thiamin	0.037 mg
Riboflavin	0.019 mg
Niacin	0.594 mg
Pantothenic acid	0.089 mg
Vitamin B6	0.08 mg
Folate, total	15 μg
Choline, total	6.7 mg
Betaine	0.1 mg
Vitamin A, RAE	42 μg
β-carotene	449 μg
α-carotene	101 μg
Lycopene	2570 μg
Lutein + zeaxanthin	123 μg
Vitamin E	0.54 mg
Vitamin K (phylloquinone)	7.9 μg
Lipids	
Fatty acids, total saturated	0.028 g
Fatty acids, total monounsaturated	0.031 g
Fatty acids, total polyunsaturated	0.083 g

(Chaudhary et al. 2018). Phenolic acid content increases during the early stages of ripening but decreases after the pink stage (Nour et al. 2014) and is negatively affected by solar UV radiation (Sharma et al. 2018). Phenolic acid, vitamin, and glycoalkaloid

contents are also affected by the type of cultivar (Chaudhary et al. 2018). Moreover, bioavailability and the levels of tomato constituents, such as lycopene, can be altered by processing such as cooking. However, genetic architecture or genotype is the major factor determining HR trait variation within species (Manganaris et al. 2018). Improvement of the nutritional and HR composition of tomato is complex because these traits are under polygenic control, have quantitative inheritance, and are affected by environmental conditions (Fernie et al. 2006).

1.3 Importance in the Face of Chronic Diseases and Malnutrition

The world's human population is predicted to reach more than nine billion within 30 years (Gerten et al. 2020). Rapid population growth causes hunger and malnutrition. According to a 2019 FAO report on food security, one in every ten people suffers from malnutrition (FAO 2019). In order to cope with malnutrition, it is not enough to increase agricultural production. The nutritional content of crops and dietary diversity must also be improved (Manganaris et al. 2018). These factors also affect disease incidence because it is well known that malnutrition and diseases (e.g., cancer, cardiovascular problems, blindness) are closely related (Baldermann et al. 2016). Poor nutrition can also result in increased body mass index which is a causal factor of diseases like obesity and type 2 diabetes (Trujillo et al. 2006). Therefore, a balanced, diverse, and healthy diet may prevent many diseases.

Examination of the effects of nutrition and diet on human health has gained popularity in recent years (Campestrini et al. 2019). Although it is known that nutrients affect gene expression in humans, most chronic diseases are controlled by more than one gene, and the exact mechanisms are not known (Hall et al. 2008). However, it is clear that maintaining cell and organ homeostasis is related to maintaining human health. Bioactive compounds support human health by helping to maintain this homeostasis. Plants are especially valuable sources of metabolites and contain more than 200,000 compounds with huge structural diversity (Hall et al. 2008). Many of these metabolites are associated with the prevention of various diseases. In the last decade, the nutrient and bioactive compound content of tomato has been extensively studied. It has been discovered that tomato metabolites have various bioactivities such as antihypertensive, antioxidant, anticancer, anti-inflammatory, antidiabetic, anti-allergenic, antiatherogenic, antithrombotic, antimicrobial, vasodilator, and neuro and cardioprotective effects (Ramos-Bueno et al. 2017; Ali et al. 2020). Regular and adequate consumption of tomatoes is reported to prevent certain human diseases and increase lipid peroxidation levels (Ilahy et al. 2016).

1.4 Limitations of Conventional Breeding and Rationale for Alternative Approaches

Conventional plant breeding relies on selecting useful agronomic traits based on phenotype (Phan and Sim 2017). Germplasm is screened for the desired traits, and

selected parental lines are crossed to combine these traits in a new genotype (Ishitani et al. 2004). In this method, selection of parents and progenies is mainly according to their appearance/morphology. The progeny carries genes/traits from each parent, but their effects on phenotype may not be detected until the plant is grown and the crop harvested. The main disadvantages of conventional breeding are its intensive use of labor and time and the ineffectiveness of subjective phenotyping for many traits (Moose and Mumm 2008). Despite these disadvantages, conventional breeding has been very successful in improving yield, some quality parameters, and stress resistance in tomato (Dalal et al. 2006). In addition, compared to other approaches, conventional breeding is more acceptable to consumers because it seems more natural.

On the other hand, genotype-based selection and breeding are more precise, rapid, and cost-effective than conventional breeding (Yang et al. 2016). Next-generation sequencing (NGS) platforms provide high-throughput and cost-effective DNA sequencing which allows identifying the relationship between genotypic and phenotypic variation (Barabaschi et al. 2016). In this context, whole genome sequencing and re-sequencing techniques are being used to breed new crop cultivars that are rich in nutritional compounds (Phan and Sim 2017). Re-sequencing of different varieties helps to identify single nucleotide polymorphisms (SNPs) which enable genome-wide association studies (GWAS). Thus, quantitative trait loci (QTLs) can be identified with virtually no limitation on marker availability and can be used in marker-assisted selection (MAS) studies (Barabaschi et al. 2012, 2016; Phan and Sim 2017). MAS provides some advantages over conventional breeding methods. MAS is relatively cheap and rapid, can be applied at seedling stages, and is independent of environmental influences (Collard and Mackill 2008). Moreover, the availability of genome sequence information provides thousands of markers for genomic selection, allows analysis of genetic diversity, and enables site-specific mutagenesis by genome editing (Barabaschi et al. 2016; Phan and Sim 2017). The integration of NGS technologies and other multidisciplinary platforms should accelerate crop improvement and allow the manipulation of more complex traits.

2 Description of Nutritional Compounds

2.1 Chemical Composition, Structures, and Biochemical Pathways

Tomato is rich in nutritional and bioactive compounds. In addition to its proximate composition, the main nutritional and medicinal components found in tomato are carotenoids, phenolic acids, vitamins, polyamines, GABA, and glycoalkaloids. The composition and concentrations of those compounds in tomatoes vary depending on many factors such as the tomato variety/cultivar, part of the fruit (peel, flesh, seed), extraction process, analytical approaches, and environmental conditions (Ali et al. 2020; Lu et al. 2019) as described in Sect. 5.

2.1.1 Proximate Composition, Dietary Fiber, Minerals, and Amino Acids

The proximate constituents of tomato are protein, fat, moisture, ash, and carbohydrates. Proximates and dietary fiber are especially important for determining quality parameters and regulatory processes in the food industry (Thangaraj 2016). Proteins are important biological macromolecules with different cellular functions, and their analysis is important for nutritional labeling in foods (Ali et al. 2020). Protein is mostly found in tomato seeds with only 0.88 g/100 g serving in ripe fruit (Table 1, USDA FoodData Central 2019). Tomato also has very little total lipids and ash. Moisture content of ripe fruit is 94.5% with 3.89 g of carbohydrates/100 g serving.

Fiber consists of carbohydrates (oligosaccharides, polysaccharides, and lignins) that are digested in the large intestine. Tomato is reported to have lower crude fiber content than other popular vegetables (e.g., sweet pepper, cabbage, lettuce, spinach, and potato), with approximately 0.3 g fiber/100 g dry weight (DW) or 1.2 g/100 g fresh weight (FW) (Hanif et al. 2006; USDA FoodData Central 2019). Dietary fiber is mainly found in tomato peel (Elbadrawy and Sello 2016). According to recent reviews, the dietary fiber content of tomato peel varies between 1.32 g and 88.53 g/100 g DW (Lu et al. 2019; Ali et al. 2020).

According to USDA data, ripe tomato fruits contain ten different minerals (USDA FoodData Central, Table 1). Tomatoes are considered to be good sources of potassium, phosphorus, and calcium. They also contain 18 different amino acids including those that humans must obtain through diet. The only amino acids not present in tomato are asparagine and glutamine.

2.1.2 Carotenoids

Carotenoids are polyphenolic pigments found in the plastid membranes of leaves, flowers, and fruits. They are produced by the isoprenoid biosynthetic pathway (Martí et al. 2016). Several carotenoids are found in tomato including lycopene, β-carotene, α-carotene, γ-carotene, δ-carotene, phytoene, and lutein (Chaudhary et al. 2018). Carotenoid structures vary in terms of the presence of oxygen atoms and end group cyclization (Rao and Rao 2007). Lycopene is the immediate precursor of β-carotene with lycopene β-cyclase catalyzing this conversion. Thus, red-fruited cultivars have lower activity of this enzyme than orange-fruited ones.

Carotenoids are one of the most important compounds found in tomatoes due to their contribution to tomato's antioxidant capacity. Among carotenoids, lycopene is the most abundant carotenoid with β-carotene, also known as provitamin A, ranking second (Ilahy et al. 2017). Red tomatoes have predominantly lycopene (Kabelka et al. 2004), while orange tomatoes have mainly β-carotene (Manoharan et al. 2017). Thus, the red-fruited *S. pimpinellifolium* contains lycopene at high levels (Sharma et al. 2008), and the orange-colored wild relative, *S. cheesmaniae*, contains β-carotene at high levels (Stommel 2005). Tomato provides almost 85% of the lycopene consumed in the human diet (Rao and Rao 2007). Lycopene content increases with ripening and is found at the highest levels when ripe (Boz and Sand 2020). This increase is primarily the result of increased phytoene synthase-1 enzyme activity which catalyzes the first step of lycopene synthesis (Fraser et al. 1999).

Other carotenoids occur at low concentrations and include lutein and zeaxanthin (Raiola et al. 2014). Although the USDA database indicates that 100 g ripe tomato contains 2.57 and 0.45 mg of lycopene and β-carotene, respectively, a recent review reported that lycopene content varied from 7.8 to 18.1 mg and β-carotene content varied from 0.1 to 1.2 mg (Table 1, Martí et al. 2016). In another review, Ali et al. (2020) reported values of lycopene content from 5.0 to 11.1 mg/100 g and β-carotene content from 3.7 to 10.2 mg/100 g.

2.1.3 Phenolic Compounds

Phenolic compounds are secondary metabolites that are common in plants and are produced via the shikimate, pentose phosphate, and phenylpropanoid pathways (Lattanzio 2013; Lin et al. 2016). All phenolic compounds have at least one benzene with attached hydroxyl groups in their structure. Phenolic acids have one aromatic ring and include caffeic, ferulic, and chlorogenic acids. Polyphenols have more than one ring and include carotenoids, flavonoids, anthocyanins, and lignins.

Both phenolic acids and polyphenols make major contributions to the antioxidant capacity of tomatoes (Stratil et al. 2006). Most studies in the literature have focused on total phenolic content instead of specific phenolic acids or flavonoids. Common phenolic acids in tomato are caffeic acid, chlorogenic acid, ferulic acid, and p-coumaric acid (Ramos-Bueno et al. 2017; Sharma et al. 2018). Tomato also contains a good level of flavonoids. The most abundant flavonoids in tomato are rutin, quercetin, kaempferol, and naringenin (Ramos-Bueno et al. 2017; Sharma et al. 2018).

It is challenging to summarize the phenolic content of tomato because it varies with cultivar and environmental and physiological conditions. For instance, Martí et al. (2016) reported that 100 g of fresh tomato contains 0.9–18.2 mg naringenin chalcone, 0.5–4.5 mg rutin, 0.7–4.4 mg quercetin, 1.4–3.3 mg chlorogenic acid, 0.1–1.3 mg caffeic acid, and 0–1.3 mg naringenin. Total phenolic content was reported in the range of 21.34–31.23 mg chlorogenic acid equivalent/g extract, and total flavonoid content was reported in the range of 3.06–6.36 mg quercetin equivalent/g extract (Ali et al. 2020). Phenolic compound content is highly, positively correlated with antioxidant capacity (Stratil et al. 2006). The antioxidant capacity of tomato is difficult to compare because it changes based on the measurement method (ferric-ion reducing power, DPPH radical scavenging activity, ABTS radical cation scavenging activity) and extraction solvents and methods. Moreover, carotenoids and vitamin C and E also contribute to the antioxidant capacity of tomato (Ali et al. 2020).

2.1.4 Vitamins

The most abundant vitamins in tomato are ascorbic acid (vitamin C), provitamin A (described in Sect. 2.1.2), and vitamin E. Ascorbic acid is mainly synthesized by the L-galactose pathway in plants (Mellidou et al. 2021). It is a water-soluble antioxidant compound that neutralizes free radicals. The oxidized forms of ascorbic acids are recycled via the ascorbate-glutathione cycle. The name vitamin E encompasses four (α, β, δ, γ) tocopherols and the corresponding four tocotrienols (Raiola et al. 2015). These compounds are lipid-soluble antioxidants and only synthesized by plants. The precursors of tocopherols and tocotrienols are phytyl pyrophosphate and

geranylgeranyl pyrophosphate, respectively, which are produced in the isoprenoid pathway in plastids (Raiola et al. 2015; Mène-Saffrané 2017).

Ascorbic acid is the most significant vitamin in tomato with 13.7 mg found in 100 g fresh fruit (Table 1). During domestication, ascorbic acid levels decreased; thus, wild relatives of tomato contain higher levels of this vitamin than cultivated tomato (Mellidou et al. 2021). Ascorbic acid content is affected by various factors. Its levels increase during maturation and decrease during ripening (Chaudhary et al. 2018). Moreover, ascorbic acid levels are greatly affected by genotype. Ali et al. (2020) reported ascorbic acid levels ranging from 10.86 to 85.00 mg/100 g.

Another vitamin that contributes to the antioxidant capacity of tomato is vitamin E. The most abundant structural form of vitamin E is α-tocopherol (Raiola et al. 2015). The USDA reports the vitamin E content of tomato as 0.54 mg in a 100 g fruit. In a review, Ali et al. (2020) reported α-tocopherol levels varying from 0.59 to 0.88 mg/100 g. Folate is another nutritional compound found in tomato. Humans cannot synthesize this phytochemical, so it must be consumed in the daily diet (Vats et al. 2022). Folate content varies in the range of 4–60 μg/100 g FW (Upadhyaya et al. 2017), while in another study it was found in the range of 4.1–35.3 μg/100 g (Iniesta et al. 2009). The USDA reports folate levels of red ripe tomato as 15 μg/100 g (Table 1).

2.1.5 Polyamines

Polyamines are small aliphatic amines that occur in free, conjugated, and covalently bound forms in plants (Fortes and Agudelo-Romero 2018). The main plant polyamines are putrescine, spermidine, and spermine. Polyamine synthesis occurs in the cytoplasm. Putrescine is synthesized from the amino acid arginine and can then be converted to spermidine which can then be converted to spermine. Polyamine biosynthesis has been linked to a larger network involving abscisic acid, nitric oxide, and tricarboxylic acid cycle as well as GABA. Polyamines are important metabolites for plants due to their roles in fruit ripening and stress conditions and for humans due to their beneficial effects on health. Putrescine, spermidine, and spermine are abundant polyamines in tomato (Fortes and Agudelo-Romero 2018). Spermidine and spermine are reported to have roles in water stress tolerance (Montesinos-Pereira et al. 2014). During tomato fruit maturation, putrescine levels increase while spermidine and spermine levels decrease (Tsaniklidis et al. 2016; Gutierrez et al. 2021). Moreover, polyamine content differs based on the cultivar (Montesinos-Pereira et al. 2014).

2.1.6 Gamma-Aminobutyric Acid (GABA)

GABA is a non-proteinogenic amino acid and has been attracting attention in recent years for its health benefits. GABA is produced by the GABA shunt (Takayama and Ezura 2015). The precursor of GABA is glutamate and biosynthesis occurs in the cytoplasm. Tomato accumulates GABA at high levels. During fruit development GABA levels increase at the mature green stage but decrease during red ripe stage (Takayama and Ezura 2015). Depending on factors such as genotype, environmental conditions, and post-harvest treatment, Gramazio et al. (2020) reviewed GABA concentrations in tomato as 0.35–2.01 mg/g.

2.1.7 Glycoalkaloids

Unlike polyamines and GABA which occur throughout the plant kingdom, glycoalkaloids are nitrogen-containing compounds found in only a few species and are used in defense (Zhao et al. 2021). Solanaceous species like tomato contain steroidal glycoalkaloids (SGAs). SGAs have nitrogenous aglycone and glycoside residues and are derived from acetyl CoA via cholesterol. The two main tomato SGAs are α-tomatine and dehydrotomatine. Although they are toxic, glycoalkaloids can be used in medicine for disease treatment (Chaudhary et al. 2018). One of the important factors that affect the glycoalkaloid accumulation in tomato is fruit maturation. α-tomatine is found at high levels at green stage and decreases at the red ripe stage due to increasing enzymatic activity that metabolize the glycoalkaloids (Friedman 2002). In a review, Friedman (2002) reported that α-tomatine levels varied between 10 and 548 mg/kg at the unripe green stage and between 0.3 and 11 mg/kg at the red ripe stage in different tomato varieties.

2.2 Medicinal and Physiological Properties in Relation to Human Health

Tomato has a large part in the human diet. It is very popular due to its culinary versatility and because it is well known to contain bioactive compounds. The most abundant bioactive compounds found in tomato are carotenoids, phenolic compounds, vitamins (especially ascorbic acid), polyamines, GABA, and glycoalkaloids (Chaudhary et al. 2018). These diverse bioactive compounds play different roles in the maintenance of human health. For instance, dietary fiber has positive effects on cancer, diabetes, obesity, hyperlipidemia, and coronary heart disease (Hasegawa et al. 2017; Ali et al. 2020). Protein is essential in our daily diet because insufficient protein intake is related to risks of frailty, sarcopenia, and immunodepression (Wu 2016).

Tomato is a source of powerful antioxidants due to its rich and varied content of carotenoids, phenolic acids, ascorbic acid, and vitamin E. In addition to their antioxidant properties, these molecules have diverse bioactivity. A diet rich in carotenoids prevents certain types of cancers (Botella-Pavía and Rodriguez-Concepción 2006). Among carotenoids, lycopene has anticancer, anti-inflammatory, and immunostimulatory properties and positive effects on colitis and cardiovascular diseases (Fawzi et al. 2000; Gouranton et al. 2011; Holzapfel et al. 2013; Cheng et al. 2019). β-carotene helps in the prevention of atherosclerosis, photooxidative processes, and congestive heart disease (Stahl and Sies 2003; Karppi et al. 2013; Miller et al. 2020). It is also the precursor of vitamin A, thus preventing vitamin A deficiency (Fernández-García et al. 2012). The other minor carotenoids, lutein and zeaxanthin, protect humans against cataracts and age-related macular degeneration (Gale et al. 2003). Phenolic compounds are another category of powerful antioxidant molecules. Phenolic acids have antioxidative, antimicrobial, antiallergic, antimutagenic, anti-inflammatory, and anticancer properties (Martínez-Valverde et al. 2002; Ramos-Bueno et al. 2017). Flavonoids are a predominant type of phenolic compound in tomato and have anti-inflammatory effects and reduce cardiovascular

diseases and cancer (So et al. 1996; Tomlinson et al. 2017). Among the vitamins, ascorbic acid and vitamin E play roles in the prevention of certain cancers (Ramos-Bueno et al. 2017). Moreover, ascorbic acid has antiatherogen effects and plays role in the regulation of inflammation and insulin metabolism (Tousoulis et al. 2003). Vitamin E is important in preventing type II diabetes, cardiovascular diseases, and age-related muscular degeneration and has anti-inflammatory affects (Montonen et al. 2004; SanGiovanni et al. 2007; Cordero et al. 2010). In addition to vitamins C and E, folate also exhibits HR properties. It helps regulate the metabolism of homocysteine and reduce anemia (Solini et al. 2006; Castellanos-Sinco et al. 2015).

Polyamines, especially spermidine, also have positive effects on human health (Dala-Paula et al. 2021). Spermidine helps control blood pressure and the incidence of heart disease, has anti-aging properties, and is important for the synthesis and stabilization of nucleic acids and proteins (Eisenberg et al. 2016; Madeo et al. 2018; Muñoz-Esparza et al. 2019). GABA is another important molecule that helps to reduce blood pressure, induce relaxation, and enhance immunity (Takayama and Ezura 2015). Glycoalkaloids are important in protection against human pathogens (Chaudhary et al. 2018). Glycoalkaloids also have anticancer properties and antimicrobial activities (Iijima et al. 2013). Phytosterols play roles in the prevention of some cancer types such as colon cancer and heart disease (Ramos-Bueno et al. 2017; Uddin et al. 2018).

2.3 Methods of Nutraceutical Improvement: Agronomic and Postharvest Techniques

Plant breeding helps to improve the quality, yield, flavor and aroma, and nutritional and medicinal contents of crop plants. In addition to genotype, the levels of HR components are greatly affected by agronomic practices and environmental factors. Agronomic techniques include the regulation of UV radiation, cultivation temperature, adequate irrigation, salt stress, and cultivation methods. UV radiation (UV-B, 280–315 nm) causes DNA damage, and plants can increase the synthesis of antioxidant molecules to cope with this damage (Huché-Thélier et al. 2016). During fruit ripening, very high (>30 °C) temperatures and altered light profiles cause increased levels of both lycopene and phenolics (rutin and caffeic glucoside) (Gautier et al. 2008; Dzakovich et al. 2016). It is also reported that low (<12 °C) temperatures cause increased lycopene content (Luthria et al. 2006; Dzakovich et al. 2016). Water stress causes decreased levels of lycopene and β-carotene but increased levels of chlorogenic acid (Atkinson et al. 2011) and polyamines (Hano et al. 2017). In addition, organic pharming practices may influence the production of polyamines (Rapa et al. 2021). Salt stress may cause increased levels of phenolic acids and carotenoids due to osmoregulation (Sumalan et al. 2020).

Postharvest treatments and storage also affect the nutritional composition of tomato. After harvest, tomatoes should be stored at appropriate temperatures (4–12 °C) to maintain their nutritional quality (Chomchalow et al. 2002). Very low storage temperatures cause low levels of lycopene but high levels of putrescine

content (Tsaniklidis et al. 2021). The type of material used to coat tomatoes can also change the levels of bioactive compounds. Chitosan coating may change the anthocyanin and flavonoid content (Ortega-Ortiz et al. 2007), whereas gum arabic maintains phenolic compound and carotenoid levels (Ali et al. 2013). In addition, natural preservatives can be applied instead of chemical preservatives. Melatonin is a natural preservative and its application may induce phenolic compound synthesis (Dewanto et al. 2002).

2.4 Requirement for Biotechnological Intervention

Plants produce approximately 25,000 metabolites that significantly contribute to the human diet (Go et al. 2005). Thus, improving the nutritional quality of crops is a valuable but difficult goal of breeding strategies. Although HR traits are related to the improvement of the plant's biochemical composition, research on nutritional quality is limited (Kusano and Fukushima 2013). Moreover, genetic diversity is essential for trait improvement (Chaudhuri et al. 2022). If this diversity is not present in the crop or its relatives, the gains provided by breeding can be limited. Thus, while a mutation present in a plant's gene pool can be used to increase the content of a HR trait, the introgression of random mutations used in conventional breeding can be nonspecific and less effective than other approaches (Hartwell et al. 2018; Mao et al. 2019). Advances in genomics, transcriptomics, proteomics, and metabolomics allow the identification of biochemical pathways and expose the genetic bases of these pathways. This knowledge allows prediction of breeding value and development of selection and deployment criteria (Natalini et al. 2021). Moreover, the integration of omics technologies allows us to predict unknown gene functions (Kusano and Fukushima 2013) and understand and manipulate plant metabolism (McGloughlin 2010). Because metabolic pathways and their regulation are very complex, qualitative and quantitative profiling of metabolites using metabolomic approaches is very important. In addition, by comparison with wild relatives, this approach can reveal alleles that have been lost or silenced during domestication (Hall et al. 2008). These traits can be recovered by cross-hybridization with wild relatives or by biotechnological techniques like metabolic engineering, genome editing, and transgenesis.

Metabolic engineering and genetic manipulation methods like gene editing and transformation can be used to target specific HR traits. Metabolic engineering involves increasing the synthesis of a target compound, producing new compounds, and degrading unfavorable compounds by interfering with a key enzyme in a biochemical pathway (McGloughlin 2010). Genetically modified crops (GMOs) can also be used to overcome the disadvantages of conventional breeding. However, there are still reservations about the use of GMOs due to putative negative effects on human health and the environment (Chaudhuri et al. 2022). The major concern with GMOs is that the transgene with its viral promoter or terminator sequences remains integrated in the plant's genome often times accompanied by an antibiotic or herbicide resistance gene (Li et al. 2019). To eliminate these concerns, next-generation genome editing methods like the clustered regularly interspaced short

palindromic repeats/Cas9 (CRISPR/Cas9) system can be used. Because this genome editing tool causes mutations by inserting or deleting a small number of nucleotides within the target gene itself, these alterations are considered to be more like natural variations (Huang et al. 2016).

3 Genetic Resources of Tomato

3.1 Section *Lycopersicon*

Cultivated tomato belongs to the *Lycopersicon* section of genus *Solanum*. According to Peralta et al. (2008), wild tomato consists of 12 species in section *Lycopersicon*. These wild tomatoes include both red- and orange-fruited species: *Solanum pimpinellifolium*, *Solanum cheesmaniae* and *Solanum galapagense*. The remaining species have green or purplish-green fruits: *Solanum pennellii*, *Solanum chmielewskii*, *Solanum* chilense, *Solanum peruvianum*, *Solanum habrochaites*, *Solanum neorickii*, *Solanum arcanum*, *Solanum corneliomulleri*, and *Solanum huaylasense* (Peralta et al. 2008; Grandillo et al. 2011; Caicedo and Peralta 2013). It is possible that additional species may be added in the future as both *S. arcanum* and *S. huaylasense* were only recognized as distinct species in 2005 (Peralta et al. 2005). While all of the aforementioned species are considered crossable with *S. lycopersicum*, not all combinations result in viable seeds and fertile hybrids (Rick 1980; Díez and Nuez 2008). For instance, crosses between the red- and orange-fruited species are highly compatible; however, embryo rescue is usually required to obtain progeny from interspecific hybridization between cultivated tomato and *S. peruvianum* or *S. chilense* (Rick 1980). In addition to embryo rescue, success in interspecific hybridization has been associated with various factors including the use of *S. lycopersicum* as the female parent (Rick 1980), selection of more compatible accessions within a species (Rick 1983), and the use of pollen mixtures and bridge lines (Picó et al. 2000).

3.2 Section *Lycopersicoides*

Beyond species in section *Lycopersicon*, cultivated tomato is also crossable with *Solanum lycopersicoides* and *Solanum sitiens* in section *Lycopersicoides* (Caicedo and Peralta 2013). Both of these species have morphologies that are intermediate between tomato and potato (Knapp and Peralta 2016). Initially, hybrids of *S. lycopersicum* and *S. lycopersicoides* obtained via embryo rescue were male sterile and unilaterally incompatible if used as the female parent in backcrossing with cultivated tomato (Rick 1951). However, later work resulted in male fertile hybrids allowing introgression of *S. lycopersicoides* traits into cultivated tomato (Chetelat et al. 1997). Hybrids between cultivated tomato and *S. sitiens* were obtained using *S. lycopersicum* × *S. lycopersicoides* bridge lines (De Verna et al. 1990). Other tomato-like Solanums include *Solanum juglandifolium* and *Solanum ochranthum* in

section *Juglandifolia* which are woody and vining plants (Knapp and Peralta 2016). These species are not crossable with *S. lycopersicum* (Grandillo et al. 2011).

3.3 Germplasm Collections

Tomato germplasm has been assiduously collected, characterized, and preserved in multiple collections throughout the world. According to Genesys (https://www.genesys-pgr.org/) which consolidates data on germplasm collections, over 36,000 tomato accessions are available worldwide. The C.M. Rick Tomato Genetics Resource Center at the University of California, Davis, has approximately 4500 accessions including more than 900 wild species entries encompassing sections *Lycopersicon*, *Lycopersicoides*, and *Juglandifolia* (Chetelat 2022, https://tgrc.ucdavis.edu/). The Asian Vegetable Research and Development Center in Taiwan curates over 8500 accessions (https://genebank.worldveg.org/). The Northeast Regional Plant Introduction Station Plant Genetic Resources Unit in Geneva NY has 6600 tomato accessions representing 15 species (https://www.ars.usda.gov/northeast-area/geneva-ny/plant-genetic-resources-unit-pgru/docs/tomato-collection/). Other sizable tomato germplasm collections are found in Spain, France, Russia, the Netherlands, Australia, and Bulgaria (Díez and Nuez 2008). These collections have been and continue to be valuable resources for tomato improvement.

4 Traditional Breeding and Classical Genetics of HR Genes

Like many fruits and vegetables, the breeding of tomato has undergone different stages as defined by breeding methods and goals. The modern era of tomato breeding began during the late nineteenth century at which time landraces were available in a variety of colors indicating variable carotenoid content (Bai and Lindhout 2007; Scott et al. 2013). Since the 1940s, breeders have used wild tomatoes as a source of insect, disease, and abiotic stress resistance as well as to improve fruit quality by increasing soluble solids content (Rick 1973; Rick and Chetelat 1995). The first tomato hybrid was released in 1946 and, since then, hybrids have come to dominate the market (Bai and Lindhout 2007).

4.1 Breeding Goals

Breeding goals have shifted from a focus on yield and shelf life in the 1970s and 1980s to taste and nutrition since the 1990s. Despite this, the primary aim of most breeders is to produce a high-yielding, high-quality crop at low cost. Hence, taste and nutrition breeding will nearly always be secondary to yield and disease and abiotic stress resistance traits. Specific breeding goals also vary depending on the type of tomato. For processing tomato, the emphasis is on deep red external and internal color and high soluble solid content (4–6 °Brix). To achieve a deep red color, the lycopene

content of skin and flesh is important as are the presence and amounts of other carotenoids that alter the shade of red from yellow to orange (Díez and Nuez 2008). In the 1990s, Heinz released a series of processing cultivars with high and very high lycopene content (Scott et al. 2013). For fresh market tomato, the focus is on yield, fruit size and shape, and biotic stress resistance. In general, this has led to reduced internal fruit quality. In addition, various colors of fresh market tomato have been bred including white, yellow, orange, pink, red, purple, black, and striped (Bai and Lindhout 2007; Díez and Nuez 2008) (Fig. 1).

4.1.1 Breeding for Color

The genetic resources for color breeding include landraces and heirloom varieties as well as wild species such as *S. pimpinellifolium*, *S. habrochaites*, and *S. pennellii* for various colors and increased lycopene content (Grandillo et al. 2011). During the 1940s, the control of tomato external and internal color was described to be the result of different combinations of genes in flesh and skin (Young and MacArthur 1947). For instance, red-fruited tomatoes were described as the result of red flesh and yellow peel, while pink tomatoes have red flesh and colorless peel. In their monograph on tomato horticultural traits, Young and MacArthur (1947) listed five single genes controlling fruit color: *R* for red flesh (*r* for yellow), *t* for tangerine color, *Y* for yellow peel, m_{1-4} for modification of red color, and *G* for inhibiting the development of red color. Since then, many more carotenoid pathway and modifier genes have been identified and placed on tomato's classical linkage map (Tanksley 1993; updated in Scott et al. 2013). In 1986, Tigchelaar (1986) described seven genes that had been introgressed into tomato to breed color: *B* for high beta-carotene and its allele og_c for interior crimson color, *gs* for green stripe on ripe fruit, *hp* for high lycopene pigment, *r* for low total carotenoid content, *t* for tangerine color, and *y* for colorless peel. Since then, the repertoire of introgressed genes has expanded to include *Del*, *dps*, and *gh* for orange color; *at* for apricot color; additional *hp* genes (*hp-1* to *hp-3*) and I_p for deep red; *r* for yellow color; *ry* for red color in yellow fruit; *gf* and *Gr* for green flesh and ripe fruit; and Mo_B for modification of *B* expression (Kabelka et al. 2004; Díez and Nuez 2008). As a result of these introgressions, the carotenoid content of modern cultivars varies from deep red varieties with very high levels of lycopene to orange and yellow types with predominantly beta or delta-carotene. High pigment (*hp*) genes have been employed for enhancing dietary and HR quality because, in additional to increased lycopene and carotenoid levels, these genes are associated with improved ascorbic acid, phenolic, and flavonoid content (Levin et al. 2003). Unfortunately, such cultivars have not been widely accepted because of their reduced yield (Scott et al. 2013).

4.1.2 Breeding for Other HR Traits

Compared to carotenoid content, very little traditional breeding has addressed other nutraceutical traits. The only exception is ascorbic acid. Lincoln et al. (1943) recognized the importance of this vitamin as well as beta-carotene (provitamin A) in tomato as far back as the 1940s. Berger et al. (1966) crossed cultivated tomato with *Lycopersicon minutum* (now known as *S. chmielewskii*) and found that the hybrids had 3.5-fold more ascorbic acid than the cultivated parent. F_2 plants had highly variable concentrations

which prevented the researchers from determining the exact inheritance pattern of ascorbic acid content. This is not surprising given the quantitative nature of this trait. Variation in ascorbic acid content was also detected in more wild species (Stevens 1986), and Stevens and Rick (1986) crossed *S. peruvianum* with cultivated tomato to improve ascorbic acid content. However, to our knowledge, this work did not result in a commercial cultivar and none of the most popular processing and fresh market cultivars over the past 50 years are notable for their vitamin content (Scott et al. 2013).

As with other crops, the difficulty of studying and improving most nutraceutical traits in tomato lies in their quantitative nature. These traits are often the result of complex metabolic pathways involving multiple structural and regulatory genes. Color traits are also controlled in a complex manner; however, their visual phenotype allowed selection of single-gene mutants over thousands of years since initial domestication. Tomato's other nutraceutical components do not have the advantage of obvious phenotypes; hence, research into these compounds has only exploded with the advent of tomato molecular genetics.

5 Phenotype-Based Diversity Analysis of HR Traits

Tomato is a significant vegetable for human heath due to its high levels of crude protein and fiber in its proximate composition, as well as being rich in carotenoids, phenolic acids, vitamins, and other antioxidant compounds such as polyamines and GABA. Variation in these traits was briefly mentioned in Sect. 2.1. However, because many studies have characterized and quantified HR phytochemicals in both cultivated and wild tomato genotypes, the reported diversity in these traits is more fully described in this section.

5.1 Protein and Crude Fiber Diversity

As already mentioned, most of the protein in tomato fruits is found in the seeds. Lu et al. (2019) found that the protein content of tomato seeds varied from 23.6 to 40.9 g/100 g DW. On the other hand, the protein content of the tomato peels is much lower. For instance, Grassino et al. (2016) reported up to 1.85 g/100 g DW, while Elbadrawy and Sello (2016) found 10.50 g/100 g DW protein in tomato peel. Crude fiber is composed of cellulose, hemicellulose, and lignin. Several studies have focused on the fiber content of tomato seeds and peel. These byproducts of tomato processing can be used as ingredients in the food industry, and their high fiber content was reported by Herrera et al. (2010). Variation of fiber content in tomato germplasm was also investigated. In one study, seven tomato cultivars from the Guinea Savanna zone of South West Nigeria were characterized for fiber content, and high variation was reported with a CV (coefficient of variation) value of 133.91% (Olaniyi et al. 2010). In addition, the fiber content of three tomato cultivars (Monika, Isabella, and cherry) was investigated, and significant differences between ripening stages were reported. Cherry cultivars had the highest fiber content (Opara et al. 2012).

The most comprehensive study was performed by Al Said et al. (2014). In this study, fiber content of 54 tomato accessions grown in Oman ranged from 0.34% to 1.05%. High variation in the fiber content of 20 genotypes (0.3–1.27% range variation) was also described by Anjum et al. (2020). In addition to cultivated genotypes, wild tomato species were also evaluated for fiber content. The genetic potential of *S. pennellii* was investigated for lignin, cellulose, hemicellulose, and saccharification content using an introgression line (IL) library, and fiber content showed high variation among the tested ILs (Caruso et al. 2016). Such variation suggests that improvement of this trait in tomato fruit via breeding is possible.

5.2 Carotenoid Diversity

Lycopene and β-carotene are the most important carotenoids in tomato. Due to their significance for fruit color and high antioxidant activity, genetic variation of carotenoids was evaluated in several studies. Although these studies were performed in limited numbers of tomato genotypes (up to 53 tomato genotypes), they make it clear that genotype, ripening stage, and fruit type (green, red, and black) all affect carotenoid accumulation (Adalid et al. 2010; Roselló et al. 2011; Salunke et al. 2012; García-Valverde et al. 2013; Flores et al. 2017). Adalid et al. (2010) studied 49 accessions including local Spanish cultivated tomato, *S. lycopersicum* var. *cerasiforme* and *S. pimpinellifolium* lines. Remarkably, one of the seven *S. pimpinellifolium* accessions had ninefold the normal levels of lycopene. Roselló et al. (2011) reported that lycopene and β-carotene were affected by mainly genotype (G) effect followed by environment (E) and G × E interaction. Salunke et al. (2012) characterized 18 red and 12 orange tomato genotypes for β-carotene content and found that orange ones had the highest β-carotene content. García-Valverde et al. (2013) analyzed four fresh market tomato cultivars at different stages (green, breaker, pink, red) and five processing cultivars at the red stage for carotenoids. Tomatoes harvested at the red stage had the most lycopene and β-carotene. In addition, Flores et al. (2017) identified the effect of tomato shape (oblate to round to oblong and pyriform), color (yellow, pink, and red), and size on carotenoids. The most lycopene, β-carotene, phytoene, and phytofluene contents were measured in pink and red tomatoes, while the most lutein, violaxanthin, neoxanthin, and chlorophyll contents were found in darker-colored fruits. In another study, 69 local tomato accessions belonging to 8 cultivar groups from València (Spain) were characterized for lycopene and β-carotene as well as proximate composition traits, major sugars (glucose and fructose), and citric acid (Figàs et al. 2015). As a result, high diversity was determined among accessions. The cherry- and Penjar-type cultivars had higher values for most traits, including β-carotene dry matter, soluble solid content, titratable acidity, and taste.

5.3 Phenolic Compounds and Total Antioxidant Activity Diversity

Tomato is known for its high levels of phenolic compounds which contribute to its high antioxidant activity (Hanson et al. 2004). Flavonoids are important members of this

family of phytochemicals with a variety of different flavonoids found in various tomato cultivars (Slimestad et al. 2008; Fattore et al. 2016). Most of the studies focused on a limited number of tomato cultivars for phenolic acid composition and antioxidant traits. These studies reported significant variation in terms of phenolic acids and pointed out that the predominant phenolic acid in tomato is chlorogenic acid followed by caffeic acid, chlorogenic acid derivatives, ferulic acid, kaempferol-3-O-rutinoside, rutin, and p-coumaric acid (Martínez-Valverde et al. 2002; Minoggio et al. 2003; Spencer et al. 2005; Luthria et al. 2006; Barros et al. 2012; García-Valverde et al. 2013; Asensio et al. 2019). The most comprehensive study was performed by Bhandari et al. (2016) using 119 cherry and non-cherry genotypes including both cultivars and germplasm materials. The study revealed that total phenolic and antioxidant activity had lower variation than carotenoid, ascorbic acid, and flavonoid content. According to Figàs et al. (2015), cherry and Penjar tomatoes had higher total antioxidant and phenolic content than other types of tomatoes grown in València, Spain. In another study, 54 genotypes (cultivars, commercial hybrids, cherry tomatoes, wild species, interspecific hybrids, and backcross lines) were screened for antioxidant levels and quality. Wild species and interspecific hybrids between LA1777 (*S. habrochaites*) and an elite genotype were found to have the highest antioxidant capacity, as well as high phenolic and flavonoid contents (Kavitha et al. 2014). Other studies also highlighted the genetic potential of wild tomato species as resources for the improvement of antioxidant capacity and phenolic compounds (Kavitha et al. 2014; Top et al. 2014). A study performed by Top et al. (2014) reported that the wild tomato species *S. pimpinellifolium* (LA1589), *S. habrochaites* (LA1223), and *S. peruvianum* (LA2172) generally had higher content than cultivated tomato for water-soluble antioxidant activity and total phenolic content. Among the wild species, *S. peruvianum* had the highest values for these traits.

5.4 Vitamin Diversity

Tomato is an important dietary source of ascorbic acid. This vitamin's content is highly variable in tomato (Bhandari et al. 2016) with this variation mainly attributed to genotypic rather than environmental factors (Roselló et al. 2011). This observation was reinforced by Martí et al. (2018) who found that genotype was more important than farming method (conventional vs. organic) for the accumulation of HR compounds like ascorbic acid. Ilahy et al. (2017) reported an ascorbic acid level of 20 mg/100 g FW, and Mellidou et al. (2021) measured values from 6 to 23 mg/100 g FW in different cultivars. Among cultivated tomato types, cherry tomatoes are reported to have higher ascorbic acid content (Adalid et al. 2010; Figàs et al. 2015). For example, Adalid et al. (2010) measured ascorbic acid content in 28 *S. lycopersicum* var. *cerasiforme* accessions and identified two lines with three times the levels found in control lines. The wild tomato species *S. pimpinellifolium*, *S. habrochaites*, and *S. peruvianum* also have more ascorbic acid than cultivated tomato with *S. pimpinellifolium* (LA1589) having the highest ascorbic acid content among these three species (Top et al. 2014). In addition to the above studies, much research on ascorbic acid content in tomato has examined the effect of processing on this vitamin as well as other HR compounds (e.g., Chanforan et al. 2012; Koh et al. 2012; Badin et al. 2021).

Vitamin E content is affected by many factors. Vitamin E levels increase during ripening (Dumas et al. 2003), by the use of rainwater instead of irrigation (Pék et al. 2014), under high light intensity, at low temperature, and under salinity stress (Skłodowska et al. 2009). Vitamin E content also varies based on the tomato variety (Lenucci et al. 2006; Marsić et al. 2010). For instance, Marsić et al. (2010) recorded twofold variation in different cultivars grown in different locations. Frusciante et al. (2007) reported an α-tocopherol level of 0.62 mg/100 g FW.

5.5 Polyamine Diversity

Polyamine content is a trait which has been the focus of limited work in terms of germplasm characterization in tomato. In early work using a single line, putrescine was found to be the major polyamine at early stages of fruit development with much less spermidine and spermine (Morilla et al. 1996). The content of all three polyamines decreased throughout development and ripening. Yahia et al. (2001) also examined a single tomato cultivar and found that while spermine content decreased with fruit maturation, spermidine content remained unchanged. In the most recent and comprehensive work, Gutierrez et al. (2021) analyzed the polyamine content of eight genotypes from the green to red ripe stage. Different genotypes had significantly different polyamine contents with two Bolivian lines identified as having the highest levels at the green and red stages. No clear correlation was observed between the content of individual polyamines and maturation stage (Tsaniklidis et al. 2016; Gutierrez et al. 2021).

5.6 GABA Diversity

GABA is another HR compound that has been examined in diverse tomato material. San Marzano tomatoes were metabolically profiled and found to have high GABA content (Loiudice et al. 1995). Saito et al. (2008) screened 61 accessions including cultivars and 3 wild species and derived lines for GABA content. Content was found to vary greatly between years with a 13-fold difference between the lowest and highest accessions. In addition, fresh market cultivars had the highest levels compared with processing types, wild materials, and derived lines. Interestingly, GABA content is usually highest at the mature green stage and found in lesser amounts in ripe fruit (Akihiro et al. 2008; Saito et al. 2008; Takayama and Ezura 2015). In addition, its levels vary depending on the variety, environmental conditions, under stress conditions, and during post-harvest treatments (Gramazio et al. 2020).

5.7 Glycoalkaloid Diversity

Glycoalkaloids are an important toxic secondary metabolite in tomato. For this reason and because these toxic effects can be used to control the growth of various types of

cancers, tomato germplasm has been screened for this HR trait. In early work, Leonardi et al. (2000) examined glycoalkaloid and antioxidant content in cherry, cluster, salad, and elongated types of tomato. The elongated line had approximately fourfold more total glycoalkaloids than the other types, while cherry tomatoes had the highest antioxidant levels. Kozukue et al. (2004) analyzed glycoalkaloids in three normal and two mini-tomato cultivars from Japan. Glycoalkaloid content varied with genotype and in different tissues. On average, 11-fold more α-tomatine than dehydrotomatine was present in green fruits; however, neither compound was detectable in ripe fruits. Choi et al. (2010) monitored glycoalkaloid content during different growth stages and also found that they decreased significantly during ripening. It was reported that the α-tomatine level was 500 mg/kg FW in green fruits and then decreased to 5 mg/kg FW when the fruits were red (Friedman 2013).

In the most comprehensive work done to date, glycoalkaloids were profiled in eight accessions including cultivated tomato and the wild species *S. lycopersicum* var. *cerasiforme* (LA2213), *S. pimpinellifolium* (LA1589), *S. pennellii* (LA716), *S. habrochaites* (LA1777), *S. neorickii* (LA2133), and *S. cheesmaniae* (LA1414) (Iijima et al. 2013). A total of 123 different glycoalkaloids were found in the different accessions with distinct profiles in some of the lines. The highest levels of total glycoalkaloids were measured in *S. lycopersicum* var. *cerasiforme* and *S. cheesmaniae.* Schwahn et al. (2014) examined four of the same wild accessions as Iijima et al. (2013) but also included *S. chmielewskii* (LA1028), *S. peruvianum* (LA1274), and *S. cheesmaniae* (LA1414). They detected 169 different glycoalkaloid peaks with fruits of *S. peruvianum* and *S. chmielewskii* containing more different types than the other species. In other work, 14 Italian landraces from three broad fruit classes, one French landrace, and three F1 hybrids were analyzed for glycoalkaloids, phenols, and amino acids. While round/elongate types had more glycoalkaloids, flattened types had more phenolic compounds and amino acids (Baldina et al. 2016). Thus, in some cases, fruit type and shape can be rough indicators of health value. However, more precise prediction of phenotype for phytochemical traits requires knowledge of plant genotype. Such knowledge can be gained using molecular markers or by genome sequencing.

5.8 Other HR Compound Diversity

In addition to the compounds described above, tomato contains many other known or potential beneficial HR phytochemicals which have received less attention for one reason or another. For example, anthocyanins have anti-inflammatory, anticancer, and antimicrobial effects (Chen et al. 2021); however, anthocyanins are only produced at significant levels in fruits of a few tomato cultivars and genetically modified lines. There are also not many studies evaluating the phytosterol content of tomato cultivars. In one study, the phytosterol content of tomato seed oil (a waste product of sauce processing) extracted from three Italian processing tomato varieties (Principe Borghese, Rebelion F1 and San Marzano) was investigated. Principe Borghese had the most β-sitosterol content (Giuffrè and Capocasale 2016a). The same cultivars

were also characterized for fatty acid composition, and a study revealed their high variation in seed oil (Giuffrè and Capocasale 2016b).

5.9 Molecular Diversity

In addition to phytochemical diversity, the genetic diversity of tomato has been extensively investigated using molecular markers. Twenty years ago, such studies were performed mainly with RAPD (random-amplified polymorphic DNA) markers (Archak et al. 2002; Carelli et al. 2006; Huh et al. 2011). Germplasm characterization then began to encompass other molecular marker systems such as AFLP (amplified fragment length polymorphism) and SSR (simple sequence repeat) markers (Park et al. 2004; Cebolla-Cornejo et al. 2013; Korir et al. 2014; Zhou et al. 2015). Several genetic linkage maps derived mainly from related species such as *S. pennellii*, *S. pimpinellifolium, S. habrochaites*, and *S. chmielewskii* were constructed using molecular markers. All genetic and physical maps can be found on the Sol Genomics Network (https://solgenomics.net/cview/index.pl).

Next-generation sequencing of tomato germplasm has revealed a more complete picture of genome variation. From this perspective, a pan genome of tomato was developed by sequencing 725 phylogenetically and geographically representative accessions (Gao et al. 2019). The pan genome project revealed the absence of 4873 genes in the tomato reference genome and the negative selection of genes during tomato domestication. Structural variant data for 100 diverse tomato genotypes, assemblies, and annotations for 14 of these genotypes are available on the Sol Genomics Network (https://solgenomics.net/projects/tomato100/). The pan genome of tomato will be an invaluable tool for understanding the genetic diversity of health-related traits in this species.

6 Brief Account of Molecular Mapping of HR Genes and QTLs

Once tomato genetic resources for health-related compounds are evaluated, it is then possible to determine the molecular mechanism of their synthesis. Because HR compounds are quantitative traits, they are controlled by many loci in the genome and often with significant environmental control. This section comprises the identification of the single genes and QTLs that control HR using genomic and transcriptomic approaches.

6.1 Genomic Approaches

6.1.1 Biparental Mapping

Early studies focused on the identification of genes and QTLs controlling lycopene and β-carotene content in tomato as discussed in Sect. 4. Other studies in tomato have focused on fruit quality traits such as phenolic compounds, flavonoids, and total antioxidant activity. The first study was performed by Rousseaux et al. (2005).

In this work, ascorbic acid, phenolics, lycopene, and β-carotene contents and water-soluble antioxidant activity were measured in fruits from an IL library developed using *S. pennellii.* A total of 20 QTLs controlling the related traits were identified. The same *S. pennellii* IL library was used to map QTLs controlling the antioxidants carotenoids and tocopherols as well as molecular signatures for 2000 compounds. Introgressions from chromosomes T3, T6, T8, and T12 were associated with carotenoid and tocopherol content (Perez-Fons et al. 2014). In other research, a BC2F2 population derived from *S. habrochaites* was used to map the QTLs for total water-soluble antioxidant capacity, ascorbic acid, total phenolics, flavonoids, and lycopene contents. As a result, 48 QTLs were mapped (Ökmen et al. 2011). In terms of antioxidant traits, the most comprehensive study was performed by Çolak et al. (2020). In this study, an IBL (Inbred Backcross Line) population derived from *S. pimpinellifolium* was genotyped with GBS (genotyping by sequencing)-based SNP (single-nucleotide polymorphism) markers, and QTL mapping was performed for carotenoids, ascorbic acid, vitamin E, glutathione, and phenolic acid contents. As a result, 64 QTLs were detected with candidate genes identified for several of the loci. In addition, a RIL (recombinant inbred line) population developed from *S. pimpinellifolium* was used to map QTLs for soluble solids, acidity, sugars, organic acids, ascorbic acid, and carotenoid contents. Several QTLs were mapped and, most importantly, a fruit size QTL (*frw7.1*) was found to be co-localized with QTLs controlling soluble solids, ascorbic acid, and glucose contents, dry weight/fresh weight, and a sucrose phosphate synthase gene. Moreover, two candidate genes (*1-deoxy-D-xylulose 5-phosphate synthase* and *tocopherol cyclase*) were identified for β-carotene and ascorbic acid contents (Capel et al. 2015).

Ascorbic acid is an important health-related compound, and several studies were performed to reveal the molecular mechanism of its synthesis in tomato. In a study performed by Zou et al. (2006a), cDNA clones of two genes [dehydroascorbate reductase (*DHAR*) and ascorbate oxidase (*AO*)] with roles in ascorbic acid synthesis were isolated and 14 genes were mapped to 15 loci using a *S. pennellii* (LA716) IL library. In another study, QTLs controlling ascorbic acid content were mapped using populations from two wild tomato species (an IL library of *S. pennellii* and an advanced backcross population of *S. habrochaites*) and a subspecies (a RIL population developed from a cross between cherry and large-fruited lines) (Stevens et al. 2007). Consensus QTLs were determined on chromosomes 2, 8, 9, 10, and 12. In a recent study, GBS and RNA-seq analysis revealed that subline R182, a *S. pennellii* IL with high ascorbic acid content, had 39 wild alleles replacing 33 alleles in the IL's genetic background (cv. M82). An upregulated gene for a major facilitator superfamily protein was identified in the study and is a promising candidate for increased ascorbic acid content.

In addition to nutritional compounds that are associated with good health, tomato contains secondary metabolites like glycoalkaloids which have anticancer properties. A recent study identified QTLs controlling glycoalkaloid and flavonol content using a *S. pennellii* IL library. As a result, a total of 338 QTLs were detected. Two QTLs for flavonols and one for steroidal glycoalkaloids were confirmed using IBL populations. Cross-validation revealed that the steroidal glycoalkaloid QTL was co-localized to a glycoalkaloid gene cluster on chromosome 7 in tomato (Alseekh et al. 2020).

6.1.2 Association Mapping

Association mapping in tomato for HR traits has been more limited than biparental population QTL mapping. The most comprehensive study was performed by Ruggieri et al. (2014). In this work, 96 tomato lines were genotyped with 7720 SNP markers and used as an association panel for identification of QTLs controlling ascorbic acid, β-carotene, trans-lycopene, cis-lycopene, and phenolic contents. As a result, while two markers were linked to phenolics, three markers were associated with ascorbic acid, β-carotene, and trans-lycopene. In another study, an association panel containing 174 tomato lines was genotyped using SSR markers (Zhang et al. 2016). As a result, 4 and 15 markers were found to be associated with lycopene and β-carotene contents, respectively.

6.2 Transcriptomic Approaches

In addition to genomic research such as gene mapping, analysis of the transcriptome can be used to reveal the mechanisms controlling HR traits. For example, a transcriptomic study was performed for identification of genes with roles in phenolic compound accumulation (Di Matteo et al. 2013). In this study, a microarray hybridization method was used to identify DEGs (differentially expressed genes) between M82 and *S. pennellii* IL 7-3, which has a higher level of phenolic accumulation. As a result, a MATE-transporter, a vacuolar sorting protein, and three GSTs were found to be upregulated, and two ethylene responsive factors (*ERF1* and *ERF4*) were found to be associated with phenolic accumulation. In addition, a TILLING (Targeting Induced Local Lesions IN Genome) assay revealed a mutation in the *ERF1* gene which reduced the phenolic content (Di Matteo et al. 2013). Another transcriptomic study revealed the positive role of abscisic acid (ABA) in carotenoid biosynthesis (Mou et al. 2015). The most recent transcriptomic study to identify DEGs controlling ascorbic acid and phenolics synthesis was performed by Sacco et al. (2019). In this study, transcriptomic changes during the ripening of two tomato cultivars were investigated using a RNA-seq approach. This study reported that pectin methylesterase and L-ascorbate oxidase played roles in ascorbic acid synthesis and that the chalcone synthase (*CHS*) gene had a role in phenolic accumulation. In other works, comparative transcriptome analysis of tomato and potato glycoalkaloid content allowed the identification of ten genes involved in biosynthesis (Itkin et al. 2013). These genes are located into two clusters on chromosomes 7 and 12 in tomato. Comprehensive profiling of glycoalkaloids in different tissues of eight different species also allowed refinement of the biosynthetic network that controls their production in tomato (Schwahn et al. 2014).

6.3 Limitations in Studies to Date

While many studies have mapped loci and explored the differential expression of genes controlling carotenoid, ascorbic acid, phenolic, and flavonoid contents, such work has not yet been performed for other health-related traits such as fiber, phytosterol, and fatty acid composition. Moreover, although a significant number of QTL mapping studies

have been done for certain HR traits in tomato, to our knowledge, none of the major or stable QTLs have been employed as breeding tools. High-resolution QTL mapping approaches such as association mapping and candidate gene mapping targeting pathway and regulatory genes should be performed to elucidate the molecular genetic mechanism of nutritional compound accumulation. Such approaches can accelerate the identification of functional SNPs that can be used in selection to increase nutritional compound content via breeding. Moreover, it must be noted that utilization of the identified QTLs in breeding can be difficult due to the effects of genome background and environment on phenotype. As a result, genomic selection models should be developed for efficient selection of genotypes with high HR compound content without the need for extensive and expensive metabolite analysis (Cappetta et al. 2020).

7 Marker-Assisted Breeding for HR Traits

As previously mentioned, health-related traits in tomato are quantitative traits and affected by the environment (Schauer et al. 2008). As a result, breeding for HR characteristics is difficult. While several QTLs controlling HR traits have been mapped in the tomato genome as described in Sect. 6, the utilization of these QTLs in breeding depends on their stability in different environments and genomic backgrounds. When stable loci are identified, linked markers can be used in MAS (marker-assisted selection) for the trait of interest (Chaïb et al. 2006). To our knowledge, the first marker-assisted breeding strategy for HR traits was applied by Lecomte et al. (2004). In this work, five previously identified QTLs for organoleptic quality (soluble solids, reducing sugar, acidity, and pH), originating from a cherry tomato (Cervil) with good taste and aroma attributes, were introgressed to a fresh market recipient parent using a marker-assisted backcross (MABC) strategy. As a result, three improved BC3S3 lines were developed. In another study, two introgression lines (IL7-3 and IL12-4) developed from *S. pennellii* with QTLs for increased ascorbic acid, phenols, and soluble solids were crossed to the recurrent parent M82 (Sacco et al. 2013). In addition, F1 hybrids were developed by crossing two ILs. Molecular marker analysis of the F2 lines from these hybrids allowed identification of four individuals that carried the two introgressions in the homozygous state. Selfed F3 lines had high contents of ascorbic acid, phenols, and soluble solids. Moreover, fruit extracts of these lines and ILs (IL7-3 and IL12-4) showed selective cytotoxic effects on cancer cells (Rigano et al. 2014). Despite these significant studies, utilization of MAS for HR traits is in its infancy, and more studies are needed to identify stable QTLs and allow better integration of molecular marker technologies in breeding.

8 Map-Based Cloning of HR Genes/QTLs

Tomato is a model organism for the implementation of map-based cloning due to the abundance of genome mapping studies. Despite these studies, the genes cloned in tomato are mainly related to disease resistance (reviewed by Foolad 2007). However, several studies have also been performed for fine mapping and cloning of HR genes.

8.1 Carotenoid Genes

Early studies for fine mapping of genes with agronomic importance used subNILs from wild species near-isogenic lines (NILs) to pinpoint the locations of QTLs (Monforte and Tanksley 2000; Frary et al. 2003; Yates et al. 2004). Of this early research, only Yates et al. (2004) analyzed subNILs for HR traits such as lycopene. In this work, overlapping introgressions for portions of chromosome 4 from *S. peruvianum* and *S. hirsutum* were examined. As a result, a subNIL containing a 3 cM introgression was found to increase lycopene content. Also, a 4 cM introgression increased organic acid, total titratable acid, and total sugar contents.

Although the generation of subNILs increases the resolution of QTL mapping, it is inefficient for gene cloning. Thus, map-based cloning strategies were developed for positional cloning of genes. This approach was initially used for cloning genes for color traits such as carotenoid content. In early work, the *LYCOPENE* δ-*CYCLASE* (*CRTL-E*) gene that changes lycopene into δ-carotene and leads to orange color in the *Delta* mutant was cloned using a cDNA sequence from *Arabidopsis thaliana* as a probe in a cDNA library of tomato. The *CRTL-E* gene co-segregated with *DELTA* on chromosome 12 (Ronen et al. 1999). Another study focused on the *BETA (B)* gene for increased β-carotene content and *OLD-GOLD* (*OG*) which decreases β-carotene and increases lycopene content (Ronen et al. 2000). In this work, a yeast artificial chromosome (YAC) containing the region of interest was identified and screened with the markers TM16 and TG275. It was found that the *BETA* gene encodes lycopene β-cyclase which converts lycopene to β-carotene and that *og* is an allele of *BETA*. In another study, the *tangerine* locus which causes accumulation of pro-lycopene instead of all-trans-lycopene was cloned using bacterial artificial chromosomes (BACs) after identification of a co-segregating marker in an F2 population (Isaacson et al. 2002). The study showed that *tangerine*, re-designated *CRTISO*, encodes a carotenoid isomerase with a function in carotenoid desaturation. Another single recessive gene, *YFT1*, causing yellow tomato fruits was cloned after fine mapping to an 88.2 kb interval on chromosome 9 in an F2 population (Gao et al. 2016). Among the 12 identified candidate genes in the region, *Solyc09g007870*, annotated as encoding the ETHYLENE INSENSITIVE2 (EIN2) protein, was the most likely candidate gene. The study also showed that reduced expression of the *EIN2* gene in the *yft1* mutant resulted in abnormal carotenoid production. Another single recessive gene, *YFT2*, associated with yellow-fruited cherry tomato on chromosome 3 was cloned by Chen et al. (2019). In this work, fine mapping of the locus was performed using 1898 F2 plants and 10 cleaved amplified polymorphic sequence (CAPS) markers. The locus was mapped to a target interval of 95.7 kb, and annotation of the interval revealed 11 candidate genes with *PSY1* (*Solyc03g031860*), annotated as a phytoene synthase (*PSY1*), as the most likely candidate. Overexpression of this fruit-specific gene increased carotenoid content and resulted in red fruit.

8.2 Ascorbic Acid Genes

In addition to cloning genes involved in carotenoid synthesis, several studies have cloned genes for ascorbic acid synthesis. Firstly, a GDP-D-mannose pyrophosphorylase

(*GMP*) gene involved in ascorbic acid biosynthesis was cloned in tomato using the potato GMP cDNA sequence as a query against the tomato EST database (Zou et al. 2006b). As a result, a total of 65 homologous ESTs were identified in tomato. Assembly of the ESTs and RACE-PCR (rapid amplification of cDNA ends) revealed the full 1086 bp *GMP* gene sequence. The RACE-PCR method was also used to clone dehydroascorbate reductase (*DHAR1* and *DHAR1*) and ascorbate oxidase (*AO*) genes (Zou et al. 2006a).

8.3 Glycoalkaloid Genes

Glycoalkaloid content of tomato fruit affects their palatability. As a result, this trait has been the subject of research to identify the genes responsible for fruit bitterness. Fine mapping and resequencing of 650 tomato accessions showed that the *GORKY* gene is responsible for exporting α-tomatine from the vacuole to the cytosol in tomato (Kazachkova et al. 2021). Once in the cytosol, the α-tomatine is converted to non-bitter compounds. Bitter tomato genotypes were discovered to have deletions in this gene.

9 Editing of HR Traits

As explained in a Sect. 7 of this chapter, although several genes and QTLs controlling HR traits have been mapped in the tomato genome, integration of these loci in tomato classical and molecular breeding strategies is limited. Thus, alternative approaches were developed for efficient breeding of tomato cultivars with improved HR traits. Genome editing tools such as RNA interference (RNAi), insertional mutagenesis, and CRISPR/Cas9 editing provide great opportunities, not only for verification of candidate genes but also for improvement of cultivars for various traits. CRISPR/Cas9 is especially popular in crop plants including tomato. Genome editing has been successfully used in tomato for increasing abiotic and biotic stress resistance (reviewed by Chandrasekaran et al. 2021; Salava et al. 2021; Xia et al. 2021). In addition, it has been used for improvement of HR traits in several studies.

9.1 Carotenoid Content

Color traits such as lycopene and carotenoid content are major targets of genome editing. In early work, downregulation of the photomorphogenesis regulatory gene, *DET1*, using RNAi gene editing technology increased carotenoid and flavonoid content without affecting other fruit quality traits (Davuluri et al. 2005). In work by Sun et al. (2012), RNAi-mediated suppression of the *SlNCED1* gene, which has a role in ABA biosynthesis, provided higher levels of lycopene and β-carotene.

In addition to RNAi, CRISPR/Cas9 technology has also been used to modify HR content in tomato. Gene editing of *LEAFY-COTYLEDON1-LIKE4* (*L1L4*), a transcription factor with functions in development, seed storage, and protein and fatty

acid synthesis, was performed with a ZFN-mediated genome editing tool (Hilioti et al. 2016; Gago et al. 2017). Editing resulted in increased β-carotene content as well as enhanced antioxidant activity. In a study aimed at demonstrating the efficiency of the system in tomato, two carotenoid-related genes, *PSY1* and *CRTR-B2,* were successfully edited using CRISPR/Cas9 (D'Ambrosio et al. 2018). Li et al. (2018a) targeted five genes (*SGR1*, *LCY-E*, *BLC*, *LCY-B1*, and *LCY-B2*) involved in the metabolic pathway for converting lycopene to other carotenoids. The expression of these genes was knocked down by the CRISPR/Cas9 system. As a result, lycopene content increased by 5.1-fold. In other works, four genes, *SELF-PRUNING (SP), OVATE (O), FRUIT WEIGHT 2.2 (FW2.2)*, and *LYCOPENE BETA CYCLASE (CYCB)*, were edited in the wild tomato *S. pimpinellifolium* (Zsögön et al. 2018). These genes were selected to mimic the process of domestication. Editing resulted in increased fruit size and number and improved lycopene level by fivefold.

9.2 Vitamin Content

In early work, RNAi was used for downregulation of a mitochondrial APX (*mitAPX*) gene in tomato to increase ascorbic acid content (Zhang et al. 2011a). Gene editing of the *L1L4* transcription factor was also associated with higher levels of ascorbic acid (Hilioti et al. 2016; Gago et al. 2017). In perhaps one of the most exciting developments in genome editing, Li et al. (2022) knocked out the activity of a *7-DEHYDROCHOLESTEROL REDUCTASE* isoform (*Sl7-DR2)* in tomato. This enzyme normally converts 7-DHC to cholesterol preparatory to α-tomatine synthesis. By modifying phytosterol biosynthesis in this way, the researchers induced the accumulation of provitamin D3 in ripe fruits with no negative effects on growth or yield. It is important to note that ripe tomato normally contains no provitamin D3. In this way, knowledge of complex, interacting metabolic pathways was leveraged to develop a new biofortified crop.

9.3 Other Compounds: GABA, Glycoalkaloid, and Anthocyanin Contents

GABA content of tomato has received considerable attention in gene editing studies. Early work by Koike et al. (2013) used RNAi to knock down expression of three GABA catabolism genes. While constitutive downregulation of these genes was associated with a sevenfold increase in ripe fruit, the plants were phenotypically abnormal. The use of a fruit-ripening promoter resulted in normal plants with a 2.5-fold increase in GABA at the red ripe stage. GABA content was also increased up to 15-fold by removing inhibitory domains in two *GLUTAMATE DECARBOXYLASE* (*GAD*) genes; however, some of the edited plants had reduced height, delayed or no flowering, and reduced fruit size and set (Nonaka et al. 2017). Li et al. (2018a) targeted five genes in the GABA pathway using the pYLCRISPR/Cas9 system.

Similar to previous work, while significant gains in GABA content were achieved, negative effects on plant phenotype were also observed.

Glycoalkaloid content in tomato hairy roots was modified by knocking out a pathway gene using CRISPR/CAS9 (Akiyama et al. 2019). Reduced *SlS5αR2* (a DET2 homolog) expression was associated with decreased accumulation of α-tomatine and increased levels of dehydrotomatine.

Anthocyanin content of tomato has also been altered using genome editing tools. The *FEEBLY* (*FB*) gene was mutated using transposon-mediated mutagenesis and anthocyanin content increased significantly (van der Biezen et al. 1996). In another study, the anthocyanin gene (*ANT1*) was overexpressed using TALENs and CRISPR/Cas9. Both methods showed similar efficiency, and overexpression of the gene resulted in higher anthocyanin content and purple tissue (Salava et al. 2021).

10 Genetic Engineering of HR Traits

In addition to genome editing studies, transgenic approaches such as overexpression of genes have been successfully used to improve HR traits in tomato. This method can use genes that are already found in tomato (cisgenesis) or those from other species (transgenesis). In both cases, the resulting transgenic individuals are considered to be genetically modified organisms (GMOs) and, therefore, subject to regulatory control.

10.1 Carotenoid Content

The most popular HR trait in transgenic studies is carotenoid content. Early studies focused on the role of the *BETA-CYCLASE* (*β-LCY*) gene in the carotenoid pathway. The first study transformed this gene into tomato using both overexpression (OE) and antisense (AS) constructs (Rosati et al. 2000). While overexpression led to higher levels of β-carotene, the antisense construct decreased *β*-carotene content. Fruit-specific expression of lycopene *B-CYCLASE* (*β-LCY*) and *L-CAROTENE HYDROXYLASE* (*β*-CHY) genes resulted in increases in *β*-carotene, β-cryptoxanthin, and zeaxanthin contents (Dharmapuri et al. 2002). In a similar work, overexpression of lycopene *β-LCY* in tomato verified the importance of this gene by increasing β-carotene content 4.1-fold in transgenic tomato (Guo et al. 2012). In another study, plastid expression of the *β-LCY* gene improved content of *β*-carotene and provitamin A by fourfold (Wurbs et al. 2007). Apel and Bock (2009) transferred *β-LCY* genes from *Erwinia herbicola* and *Narcissus pseudonarcissus* to the plastid genome. Although the eubacterial enzyme did not affect carotenoid accumulation, the plant enzyme increased *β*-carotene and provitamin A levels in fruits. In a more recent study, overexpression of the *β-LCY* gene increased β-carotene content and was associated with increased ABA content and extended shelf life (Diretto et al. 2020). Some of the previously developed transgenic lines overexpressing lycopene *β-LCY* (from studies by D'Ambrosio et al. 2004; Wurbs et al. 2007; Apel and Bock 2009) were further tested for their metabolic profiles (Mi et al. 2022). The transgenic lines had altered

hormone content including changes in ABA, gibberellins, cytokinins, auxin, and jasmonic acid levels. The plants also had enhanced xanthophyll and provitamin A contents and increased abiotic stress tolerance.

In addition to *LYCOPENE β-CYCLASE*, *PHYTOENE SYNTHASE* and *TAXADIENE SYNTHASE* genes isolated from the bacterium *Erwinia uredovora* and the yew tree *Taxus baccata*, respectively, were overexpressed in tomato. The transgenic plants had increased total carotenoid content and also contained taxadiene (Fraser et al. 2002; Kovacs et al. 2007). Taxadiene is usually not synthesized by tomato but is valuable as a precursor of taxol, an important anticancer agent.

Transcription factors have also been used to increase carotenoid content in transgenic studies. In one study, upregulation of a *R2R3-MYB* transcription factor (*VVMYB5B*) in tomato reduced phenylpropanoid metabolism and increased the β-carotene content in tomato (Mahjoub et al. 2009). Overexpression of a brassinosteroid (*BR*) (response transcription factor) in tomato also enhanced carotenoid content and resulted in improved soluble solid, soluble sugar, and ascorbic acid levels (Liu et al. 2014). Overexpression of *SlPRE2*, a *bHLH* transcription factor, reduced chlorophyll and carotenoid abundance in tomato (Zhu et al. 2017). In addition, overexpression of the non-heme di-iron carotene betahydroxylase-encoding gene (*CRTR-B2*) increased xanthophyll content in tissues containing chloroplasts and chromoplasts (D'Ambrosio et al. 2011). In addition to transcription factors, the positive effects of plastid chaperones such as *ORANGE* (*OR*) and *HSP70* on carotenoid accumulation has been demonstrated (D'Andrea and Rodriguez-Concepcion 2019; Yazdani et al. 2019).

10.2 Flavonoid Content

Transgenic approaches have also been implemented to increase flavonoid content in tomato. The first study demonstrated that overexpression of the biosynthetic pathway genes *CHALCONE ISOMERASE*, *CHALCONE SYNTHASE*, and *FLAVONOL SYNTHASE* increased total fruit flavonols (Verhoeyen et al. 2002). In another study, the introduction of structural flavonoid genes (*STILBENE SYNTHASE*, *CHALCONE SYNTHASE*, *CHALCONE REDUCTASE*, *CHALCONE ISOMERASE*, and *FLAVONE SYNTHASE*) produced fruit peels with increased phytochemical content including flavonols and flavones (Schijlen et al. 2006). Transgenic approaches have also revealed the role of transcription factors in controlling the accumulation of flavonoid compounds. In one study, overexpression of the maize transcription factor genes *LC* and *C1* resulted in increased flavonols (Bovy et al. 2002). *AtMYB12* originating from *Arabidopsis* was also upregulated in tomato and was associated with increased levels of flavonols and chlorogenic acid in both leaf and fruit (Pandey et al. 2015).

10.3 Ascorbic Acid Content

The transgenic approach has also been implemented for improvement of ascorbic acid content in tomato. In one study, transgenic plants overexpressing the *SlGME1 and*

SlGME2 genes from the ascorbic acid biosynthesis pathway increased total ascorbic acid in leaves and fruits (Zhang et al. 2011b). Overexpression of two genes, *MONODEHYDROASCORBATE REDUCTASE* (*MDHAR*) and *DEHYDROASCORBATE REDUCTASE* (*DHAR*), encoding ascorbate recycling enzymes was performed by Haroldsen et al. (2011). In this work, overexpression of *DHAR* increased ascorbic acid content, but *MDHAR* overexpression was associated with reduced content. In a similar research, overexpression of a *DHAR* gene increased DHAR activity and ascorbic acid content in both leaves and fruits (Li et al. 2012). Upregulation of the *L-GULONO-Γ-LACTONE OXIDASE* (*GLOase*) gene boosted ascorbic acid content by 1.5-fold and also improved abiotic stress resistance (Lim et al. 2012).

In some cases, an increased level of a HR compound is a benefit of engineering aimed at another metabolite. For example, a *STILBENE SYNTHASE* cDNA involved in resveratrol synthesis and isolated from grape (*Vitis vinifera*) was overexpressed in tomato (Giovinazzo et al. 2005). As a result, transgenic fruit accumulated trans-resveratrol and had increased levels of ascorbate and glutathione, another water-soluble antioxidant.

10.4 Polyamine Content

Polyamines were an early target of efforts to increase their content via transformation methods. In 2002, Mehta et al. introduced a yeast *S-ADENOSYLMETHIONINE DECARBOXYLASE* (*SPE2*) gene with a ripening promoter into tomato. In addition to significantly higher amounts of spermidine and spermine, ripe fruit also had an unexpected threefold increase in lycopene content. A similar construct was transformed into tomato by Mattoo et al. (2007) who noted that the increased polyamine content was accompanied by reduced levels of some amino acids and sugars. Further analysis of the transgenic plants indicated that they had increased shelf life, higher levels of lycopene, and an obovoid fruit shape (Nambeesan et al. 2010). In a similar work, introduction of a human *S-ADENOSYLMETHIONINE DECARBOXYLASE* gene with a fruit-specific promoter was associated with enhanced levels of polyamines, lycopene, ascorbic acid, and soluble solids (Madhulatha et al. 2014). The use of an apple *SPERMIDINE SYNTHASE* gene driven by a constitutive promoter was equally effective at increasing both polyamine and lycopene contents in tomato (Neily et al. 2011).

10.5 GABA Content

GABA content has also been engineered in tomato. Constitutive overexpression of *SlGAD3* had fivefold more GABA than nontransgenic plants at the red ripe stage (Takayama et al. 2015). Interestingly, unlike the edited plants, the transgenic plants did not have abnormal phenotypes. In a continuation of this work, Takayama et al. (2017) co-expressed *SlGAD3* with a copy of the same gene missing the C-terminal

autoinhibitory domain. This approach yielded fruits with 18-fold higher GABA content but which had orange instead of red fruit.

10.6 Glycoalkaloid Content

Only a limited number of studies have been aimed at engineering glycoalkaloid content in tomato with much more work centered on potato. Overexpression of a *JASMONATE-RESPONSIVE ETHYLENE RESPONSE FACTOR* transcription factor (JRE4) in tomato led to a 1.8-fold increase in α-tomatine level in leaves (Nakayasu et al. 2018). Based on this work, JRE4 was determined to be a primary master regulator of glycoalkaloid synthesis. Overexpression of a *α-TOMATINE 23-HYDROXYLASE* (*Sl23DOX*) caused the accumulation of the pathway intermediates neorickiioside B and lycoperoside C in tomato leaves (Nakayasu et al. 2020). Silencing of the same gene resulted in increased α-tomatine levels. In this way, the glycoalkaloid biosynthetic pathway can be manipulated to produce a compounds of interest.

10.7 Anthocyanin Content

Anthocyanin content has been a popular target of genetic engineering in tomato due to its HR benefits and the fact that tomato normally produces very little of this pigment in its fruits. For instance, expression of two transcription factors (*DEL* and *Ros1*) isolated from snapdragon significantly increased tomato anthocyanin accumulation resulting in dark purple fruit (Butelli et al. 2008). This study also showed that cancer-susceptible *Trp53* mutant mice fed with the tomato had extended life spans. Overexpression of the same transcription factors by virus-induced gene silencing (VIGS) also yielded purple fruit due to higher anthocyanin content (Orzaez et al. 2009). In another work, introduction of *MYB* transcription factor (*ANT1)* alleles from *S. chilense* and *S. lycopersicum* resulted in increased anthocyanin pigments (Schreiber et al. 2012). When the effects of the two alleles were compared, higher pigmentation was achieved with the *S. chilense* allele. In a similar work, overexpression of a different MYB-type transcription factor (*SlMYB75*) resulted in increased anthocyanin accumulation and aroma volatiles (Jian et al. 2019). Moreover, overexpression of the *AtMYB90* gene from *Arabidopsis thaliana* in tomato increased anthocyanin accumulation in all plant organs (Li et al. 2018b). Thus, both transgenesis and cisgenesis have been successfully used to produce tomatoes high in anthocyanins.

11 Bioinformatics as a Tool

While databases such as NCBI, Ensembl Plants, and UniProt have an abundance of data related to the tomato genome and its products, several gene, genome, expression, and metabolite databases have been established specifically for tomato. These

resources and the tools they offer are important for mining of molecular and phenotypic data and integration of data from different studies. While not all of the resources described in this section contain health-related traits, they are described here for completeness. Moreover, these databases are continually updated to include information that is useful for those studying such traits in tomato.

The Sol Genomics Network houses whole genome data for many tomato varieties as well as phenotypic data and tools with which to analyze the data (Mueller et al. 2005; Fernandez-Pozo et al. 2015). The site also provides access to the Tomato Expression Atlas (https://tea.solgenomics.net/) that has both expression and anatomy viewers for the high-resolution map and various tomato tissues and organs (Fernandez-Pozo et al. 2017). The Tomato Functional Genomics Database (http://ted.bti.cornell.edu/) contains RNA-Seq data, microarray analyses, and metabolite information for various compounds primarily from wild species introgression lines (Fei et al. 2011). The Kasusa Tomato Genomics Database (KaTomicsDB, www.kazusa.or.jp/tomato/) has both nuclear and organellar genome sequences and extensive marker databases (Shirasawa and Hirakawa 2013; Kuwabara et al. 2021). The Tomato Genomics Resource Database (http://223.31.159.9/tomato2/) has information that overlaps somewhat with other databases but provides an attractive graphical format and search features that should be useful to users (Suresh et al. 2014). TOMATOMICS (unfortunately, not available from the published link at the time of writing this chapter) is another multi-omic database with DNA and mRNA sequence and gene structure, expression, and annotation data for the cultivars Micro-Tom and Heinz 1706 (Kudo et al. 2017). Public RNA-Seq data for tomato is integrated and normalized on the TomExpress platform (http://tomexpress.toulouse.inra.fr/) which also provides tools to mine these data (Zouine et al. 2017). The Co-expressed Pathways DataBase for Tomato (http://cox-path-db.kazusa.or.jp/tomato/) allows prediction of gene function based on co-expression profiles of genes with known function (Narise et al. 2017).

A metabolome database for tomato fruit was established more than 15 years ago (Moco et al. 2006). The Metabolome Tomato Database (MoTo DB, http://appliedbioinformatics.wur.nl/) encompasses data gleaned from the literature as well as that obtained from LC-MS analysis of 96 different tomato cultivars. The Plant Metabolic Network database (https://plantcyc.org/) contains biochemical pathways and metabolic data from 126 species including tomato (Hawkins et al. 2021). The Sol Genomics Network hosts an entryway to this database through the hub, SolCyc (http://solcyc.solgenomics.net/), with a family-specific, manually curated database for the nightshades, SolanaCyc (Foerster et al. 2018) and LycoCyc for cultivated tomato. A recent metabolome database (TOMATOMET, http://metabolites.in/tomato-fruits/) based on MS analysis of mature fruits of 25 cultivars is also available online (Ara et al. 2021).

Ruggieri et al. (2016) provided an excellent example of the integration of transcriptome, metabolome, and protein database information for analysis of a health-related trait. In this research, bioinformatics tools were used to elucidate the ascorbic acid metabolic network in tomato using gene expression analysis combined with data from the UniProt Knowledgebase, the Kyoto Encyclopedia of Genes and

Genomes (KEGG), and the metabolic pathway databases, MetaCyc and LycoCyc. More than 200 genes were found to play roles in the network including biosynthesis, translocation, and recycling.

12 Future Prospects

Tomato has always been at the forefront of studies in genetics, genomics, transcriptomics, proteomics, and metabolomics. As a result, a solid foundation of knowledge about the control of numerous phytochemical traits exists for this crop. Tomato breeders were early adapters of molecular-assisted selection techniques which have greatly facilitated improvement of some traits. In addition, genetic engineering and genome editing tools are now well established in tomato and allow manipulation of traits that are too complex for conventional and molecular breeding. Thus, the time is ripe for an explosion of research aimed at HR traits in tomato. Such research will be aimed not only on improvement of HR traits in fresh and processed tomato for human consumption but also on the development of cultivars for use by the pharmaceutical industry. It is hoped that such advances will help enhance human health and also be available in a fair and impartial manner for those who wish to benefit from them.

References

Adalid AM, Roselló S, Nuez F (2010) Evaluation and selection of tomato accessions (*Solanum* section *Lycopersicon*) for content of lycopene, β-carotene and ascorbic acid. J Food Compos Anal 23(6):613–618. https://doi.org/10.1016/j.jfca.2010.03.001

Akihiro T, Koike S, Tani R, Tominaga T, Watanabe S, Iijima Y, Aoki K, Shibata D, Ashihara H, Matsukura C, Akama K (2008) Biochemical mechanism on GABA accumulation during fruit development in tomato. Plant Cell Physiol 49(9):378–1389. https://doi.org/10.1093/pcp/pcn113

Akiyama R, Lee HJ, Nakayasu M, Osakabe K, Osakabe Y, Umemoto N, Saito K, Muranaka T, Sugimoto Y, Mizutani M (2019) Characterization of steroid 5α-reductase involved in α-tomatine biosynthesis in tomatoes. Plant Biotechnol J 36(4):253–263. https://doi.org/10.5511/plantbiotechnology.19.1030a

Al Said FA, Al Farsi K, Khan IA, Ali A, Khan MM, Iqbal Q (2014) Evaluation of adaptability and nutritional quality of 54 tomato accessions grown in Oman. J Food Agric Environ 12(2):40–50

Ali A, Maqbool M, Alderson PG, Zahid N (2013) Effect of gum arabic as an edible coating on antioxidant capacity of tomato (*Solanum lycopersicum* L.) fruit during storage. Postharvest Biol Technol 76:119–124. https://doi.org/10.1016/J.POSTHARVBIO.2012.09.011

Ali MY, Sina AA, Khandker SS, Neesa L, Tanvir EM, Kabir A, Khalil MI, Gan SH (2020) Nutritional composition and bioactive compounds in tomatoes and their impact on human health and disease: a review. Foods 10(1):45. https://doi.org/10.3390/foods10010045

Alseekh S, Ofner I, Liu Z, Osorio S, Vallarino J, Last RL, Fernie AR (2020) Quantitative trait loci analysis of seed-specialized metabolites reveals seed-specific flavonols and differential regulation of glycoalkaloid content in tomato. Plant J 103(6):2007–2024. https://doi.org/10.1111/tpj.14879

Anjum S, Ghafoor A, Ahmed M, Hamid A, Shah SZA, Naz RMM, Khaqan K (2020) Co-efficient of variability, correlation and path analysis of biochemical traits in tomato (*Lycopersicon esculentum* Mill) for commercial importance. Pure Appl Biol 9(1):46–55. https://doi.org/10.19045/bspab.2020.90006

Apel W, Bock R (2009) Enhancement of carotenoid biosynthesis in transplastomic tomatoes by induced lycopene-to-provitamin A conversion. Plant Physiol 151(1):59–66. https://doi.org/10.1104/pp.109.140533

Ara T, Sakurai N, Takahashi S, Waki N, Suganuma H, Aizawa K, Matsumura Y, Kawada T, Shibata D (2021) TOMATOMET: a metabolome database consists of 7118 accurate mass values detected in mature fruits of 25 tomato cultivars. Plant Direct 5(4):e00318. https://doi.org/10.1002/pld3.318

Archak S, Karihaloo JL, Jain A (2002) RAPD markers reveal narrowing genetic base of Indian tomato cultivars. Curr Sci 82:1139–1143

Asensio E, Sanvicente I, Mallor C, Menal-Puey S (2019) Spanish traditional tomato. Effects of genotype, location and agronomic conditions on the nutritional quality and evaluation of consumer preferences. Food Chem 270:452–458. https://doi.org/10.1016/j.foodchem.2018.07.131

Atkinson NJ, Dew TP, Orfila C, Urwin PE (2011) Influence of combined biotic and abiotic stress on nutritional quality parameters in tomato (*Solanum lycopersicum*). J Agric Food Chem 59(17):9673–9682. https://doi.org/10.1021/jf202081t

Badin EE, Quevedo-Leon R, Ibarz A, Ribotta PD, Lespinard AR (2021) Kinetic modeling of thermal degradation of color, lycopene, and ascorbic acid in crushed tomato. Food Bioprocess Technol 14:324–333. https://doi.org/10.1007/s11947-021-02579-1

Bai Y, Lindhout P (2007) Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? Ann Bot 100(5):1085–1094. https://doi.org/10.1093/aob/mcm150

Baldermann S, Blagojević L, Frede K, Klopsch R, Neugart S, Neumann A, Ngwene B, Norkeweit J, Schröter D, Schröter A, Schweigert FJ, Wiesner M, Schreiner M (2016) Are neglected plants the food for the future? Crit Rev Plant Sci 35:106–119. https://doi.org/10.1080/07352689.2016.1201399

Baldina S, Picarella ME, Troise AD, Pucci A, Ruggieri V, Ferracane R, Mazzucato A (2016) Metabolite profiling of Italian tomato landraces with different fruit types. Front Plant Sci 7:664. https://doi.org/10.3389/fpls.2016.00664

Barabaschi D, Guerra D, Lacrima K, Laino P, Michelotti V, Urso S, Valè G, Cattivelli L (2012) Emerging knowledge from genome sequencing of crop species. Mol Biotechnol 50(3):250–266. https://doi.org/10.1007/s12033-011-9443-1

Barabaschi D, Tondelli A, Desiderio F, Volante A, Vaccino P, Vale G, Cattivelli L (2016) Next generation breeding. Plant Sci 242:3–13. https://doi.org/10.1016/j.plantsci.2015.07.010

Barros L, Dueñas M, Pinela J, Carvalho AM, Buelga CS, Ferreira IC (2012) Characterization and quantification of phenolic compounds in four tomato (*Lycopersicon esculentum* L.) farmers' varieties in northeastern Portugal homegardens. Plant Foods Hum Nutr 67(3):229–234. https://doi.org/10.1007/s11130-012-0307-z

Berger S, Chmielewski T, Gronowska-Senger A (1966) Studies on the inheritance of high ascorbic acid level in tomatoes. Qual Plant Mater Veg 13(1):214–218

Bhandari SR, Cho MC, Lee JG (2016) Genotypic variation in carotenoid, ascorbic acid, total phenolic, and flavonoid contents, and antioxidant activity in selected tomato breeding lines. Hortic Environ Biotechnol 57(5):440–452. https://doi.org/10.1007/s13580-016-0144-3

Botella-Pavía P, Rodriguez-Conceptíon M (2006) Carotenoid biotechnology in plants for nutritionally improved foods. Physiol Plant 126(3):369–381. https://doi.org/10.1111/j.1399-3054.2006.00632.x

Bovy A, de Vos R, Kemper M, Schijlen E, Almenar Pertejo M, Muir S, van Tunen A (2002) High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes LC and C1. Plant Cell 14(10):2509–2526. https://doi.org/10.1105/tpc.004218

Boz Z, Sand CK (2020) A systematic analysis of the overall nutritional contribution of food loss and waste in tomatoes, spinach, and kidney beans as a function of processing. J Food Process Eng 43(11):e13509. https://doi.org/10.1111/jfpe.13509

Butelli E, Titta L, Giorgio M, Mock HP, Matros A, Peterek S, Martin C (2008) Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nat Biotechnol 26(11):1301–1308. https://doi.org/10.1038/nbt.1506

Caicedo A, Peralta I (2013) Basic information about tomatoes and the tomato group. In: Liedl BE, Labate JA, Stommel JR, Slade A, Kole C (eds) Genetics, genomics and breeding of tomato. CRC Press, Boca Raton, pp 1–36

Campestrini LH, Melo PS, Peres LEP, Calhelha RC, Ferreira ICFR, Alencar SM (2019) A new variety of purple tomato as a rich source of bioactive carotenoids and its potential health benefits. Heliyon 5(11):e02831. https://doi.org/10.1016/j.heliyon.2019.e02831

Campos MD, Félix MDR, Patanita M, Materatski P, Varanda C (2021) High throughput sequencing unravels tomato-pathogen interactions towards a sustainable plant breeding. Hortic Res 8(1): 171. https://doi.org/10.1038/s41438-021-00607-x

Cano A, Acosta M, Arnao MB (2003) Hydrophilic and lipophilic antioxidant activity changes during on-vine ripening of tomatoes (*Lycopersicon esculentum* Mill.). Postharvest Biol Technol 28(1):59–65. https://doi.org/10.1016/S0925-5214(02)00141-2

Capel C, Fernández del Carmen A, Alba JM, Lima-Silva V, Hernández-Gras F, Salinas M, Lozano R (2015) Wide-genome QTL mapping of fruit quality traits in a tomato RIL population derived from the wild-relative species *Solanum pimpinellifolium* L. Theor Appl Genet 128(10): 2019–2035. https://doi.org/10.1007/s00122-015-2563-4

Cappetta E, Andolfo G, Di Matteo A, Barone A, Frusciante L, Ercolano MR (2020) Accelerating tomato breeding by exploiting genomic selection approaches. Plants 9(9):1236. https://doi.org/10.3390/plants9091236

Carelli BP, Gerald LTS, Grazziotin FG, Echeverrigaray S (2006) Genetic diversity among Brazilian cultivars and landraces of tomato *Lycopersicon esculentum* Mill. revealed by RAPD markers. Genet Resour Crop Evol 53(2):395–400. https://doi.org/10.1007/s10722-004-0578-9

Caruso G, Gomez LD, Ferriello F, Andolfi A, Borgonuovo C, Evidente A, Ercolano MR (2016) Exploring tomato *Solanum pennellii* introgression lines for residual biomass and enzymatic digestibility traits. BMC Genet 17(1):1–13. https://doi.org/10.1186/s12863-016-0362-9

Castellanos-Sinco HB, Ramos-Peñafiel CO, Santoyo-Sánchez A, Collazo-Jaloma J, Martínez-Murillo C, Montaño-Figueroa E, Sinco-Ángeles A (2015) Megaloblastic anaemia: folic acid and vitamin B12 metabolism. Rev Med Hosp Gen Mex 78(3):135–143. https://doi.org/10.1016/j.hgmx.2015.07.001

Cebolla-Cornejo J, Roselló S, Nuez F (2013) Phenotypic and genetic diversity of Spanish tomato landraces. Sci Hortic 162:150–164. https://doi.org/10.1016/j.scienta.2013.07.044

Chaïb J, Lecomte L, Buret M, Causse M (2006) Stability over genetic backgrounds, generations and years of quantitative trait locus (QTLs) for organoleptic quality in tomato. Theor Appl Genet 112(5):934–944. https://doi.org/10.1007/s00122-005-0197-7

Chandrasekaran M, Boopathi T, Paramasivan M (2021) A status-quo review on CRISPR-Cas9 gene editing applications in tomato. Int J Biol Macromol 190:120–129. https://doi.org/10.1016/j.ijbiomac.2021.08.169

Chanforan C, Loonis M, Mora N, Caris-Veyrat C, Dufour C (2012) The impact of industrial processing on health-beneficial tomato microconstituents. Food Chem 134(4):1786–1795. https://doi.org/10.1016/j.foodchem.2012.03.077

Chaudhary P, Sharma A, Singh B, Nagpal AK (2018) Bioactivities of phytochemicals present in tomato. J Food Sci Technol 55(8):2833–2849. https://doi.org/10.1007/s13197-018-3221-z

Chaudhuri A, Halder K, Datta A (2022) Classification of CRISPR/Cas system and its application in tomato breeding. Theor Appl Genet 135(2):367–387. https://doi.org/10.1007/s00122-021-03984-y

Chen L, Li W, Li Y, Feng X, Du K, Wang G, Zhao L (2019) Identified trans-splicing of YELLOW-FRUITED TOMATO 2 encoding the PHYTOENE SYNTHASE 1 protein alters fruit color by map-based cloning, functional complementation and RACE. Plant Mol Biol 100(6):647–658. https://doi.org/10.1007/s11103-019-00886-y

Chen J, Xu B, Sun J, Jiang X, Bai W (2021) Anthocyanin supplement as a dietary strategy in cancer prevention and management: a comprehensive review. Crit Rev Food Sci Nut 8:1–3. https://doi.org/10.1080/10408398.2021.1913092

Cheng HM, Koutsidis G, Lodge JK, Ashor AW, Siervo M, Lara J (2019) Lycopene and tomato and risk of cardiovascular diseases: a systematic review and meta-analysis of epidemiological

evidence. Crit Rev Food Sci Nutr 59(1):141–158. https://doi.org/10.1080/10408398.2017.1362630

Chetelat R (2022) https://portal.nifa.usda.gov/web/crisprojectpages/1013392-conservation-and-utilization-of-germplasm-at-the-cm-rick-tomato-genetics-resource-center.html. Accessed 5 Aug 2022

Chetelat RT, Cisneros P, Stamova L, Rick CM (1997) A male-fertile *Lycopersicon esculentum* × *Solanum lycopersicoides* hybrid enables direct backcrossing to tomato at the diploid level. Euphytica 95(1):99–108. https://doi.org/10.1023/A:1002958030799

Choi SH, Lee SH, Kim HJ, Lee IS, Kozukue N, Levin CE, Friedman M (2010) Changes in free amino acid, phenolic, chlorophyll, carotenoid, and glycoalkaloid contents in tomatoes during 11 stages of growth and inhibition of cervical and lung human cancer cells by green tomato extracts. J Agric Food Chem 58(13):7547–7556. https://doi.org/10.1021/jf100162j

Chomchalow S, El Assi NM, Sargent SA, Brecht JK (2002) Fruit maturity and timing of ethylene treatment affect storage performance of green tomatoes at chilling and nonchilling temperatures. HortTechnology 12:271–283. https://doi.org/10.21273/HORTTECH.12.1.104

Çolak NG, Eken NT, Ülger M, Frary A, Doğanlar S (2020) Mapping of quantitative trait loci for antioxidant molecules in tomato fruit: carotenoids, vitamins C and E, glutathione and phenolic acids. Plant Sci 292:110393. https://doi.org/10.1016/j.plantsci.2019.110393

Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc Lond B Biol Sci 363:557–572. https://doi.org/10.1098/rstb.2007.2170

Cordero Z, Drogan D, Weikert C, Boeing H (2010) Vitamin E and risk of cardiovascular diseases: a review of epidemiologic and clinical trial studies. Crit Rev Food Sci Nutr 50(5):420–440. https://doi.org/10.1080/10408390802304230

D'Ambrosio C, Giorio G, Marino I, Merendino A, Petrozza A, Salfi L, Cellini F (2004) Virtually complete conversion of lycopene into β-carotene in fruits of tomato plants transformed with the tomato lycopene β-cyclase (tlcy-b) cDNA. Plant Sci 166(1):207–214. https://doi.org/10.1016/j.plantsci.2003.09.015

D'Ambrosio C, Stigliani AL, Giorio G (2011) Overexpression of CrtR-b2 (carotene beta hydroxylase 2) from *S. lycopersicum* L. differentially affects xanthophyll synthesis and accumulation in transgenic tomato plants. Transgenic Res 20(1):47–60. https://doi.org/10.1007/s11248-010-9387-4

D'Ambrosio C, Stigliani AL, Giorio G (2018) CRISPR/Cas9 editing of carotenoid genes in tomato. Transgenic Res 27(4):367–378. https://doi.org/10.1007/s11248-018-0079-9

D'Andrea L, Rodriguez-Concepcion M (2019) Manipulation of plastidial protein quality control components as a new strategy to improve carotenoid contents in tomato fruit. Front Plant Sci 10:1071. https://doi.org/10.3389/fpls.2019.01071

Dalal M, Dani RG, Kumar PA (2006) Current trends in the genetic engineering of vegetable crops. Sci Hortic 107:215–225. https://doi.org/10.1016/j.scienta.2005.10.004

Dala-Paula B, Starling M d F, Gloria MBA (2021) Vegetables consumed in Brazilian cuisine as sources of bioactive amines. Food Biosci 40:100856. https://doi.org/10.1016/j.fbio.2020.100856

Davuluri GR, Van Tuinen A, Fraser PD, Manfredonia A, Newman R, Burgess D, Bowler C (2005) Fruit-specific RNAi-mediated suppression of DET1 enhances carotenoid and flavonoid content in tomatoes. Nat Biotechnol 23(7):890–895. https://doi.org/10.1038/nbt1108

De Verna JW, Rick CM, Chetlat RT, Lanini BJ, Alpert KB (1990) Sexual hybridization of *Lycoperscion esculentum* and *Solanum rickii* by means of a sesquidiploid bridging hybrid. Proc Natl Acad Sci U S A 87:9496–9490. https://doi.org/10.1073/pnas.87.23.948

Delzenne NM, Olivares M, Neyrinck AM, Beaumont M, Kjølbæk L, Larsen TM, Benítez-Páez A, Romaní-Pérez M, Garcia-Campayo V, Bosscher D, Sanz Y, van der Kamp JW (2020) Nutritional interest of dietary fiber and prebiotics in obesity: lessons from the MyNewGut Consortium. Clin Nutr 39(2):414–424. https://doi.org/10.1016/j.clnu.2019.03.002

Dewanto V, Wu X, Adom KK, Liu RH (2002) Thermal processing enhances the nutritional value of tomatoes by increasing total antioxidant activity. J Agric Food Chem 50(10):3010–3014. https://doi.org/10.1021/jf0115589

Dharmapuri S, Rosati C, Pallara P, Aquilani R, Bouvier F, Camara B, Giuliano G (2002) Metabolic engineering of xanthophyll content in tomato fruits. FEBS Lett 519(1-3):30–34. https://doi.org/10.1016/S0014-5793(02)02699-6

Di Matteo A, Ruggieri V, Sacco A, Rigano MM, Carriero F, Bolger A, Barone A (2013) Identification of candidate genes for phenolics accumulation in tomato fruit. Plant Sci 205:87–96. https://doi.org/10.1016/j.plantsci.2013.02.001

Díez MJ, Nuez F (2008) Tomato. In: Prohens J, Nuez F (eds) Vegetables II. Springer, New York, pp 249–323

Diretto G, Frusciante S, Fabbri C, Schauer N, Busta L, Wang Z, Giuliano G (2020) Manipulation of β-carotene levels in tomato fruits results in increased ABA content and extended shelf life. Plant Biotechnol J 18(5):1185–1199. https://doi.org/10.1111/pbi.13283

Dumas Y, Dadomo M, di Lucca G, Grolier P (2003) Effects of environmental factors and agricultural techniques on antioxidant content of tomatoes. J Sci Food Agric 83:369–382. https://doi.org/10.1002/jsfa.1370

Dzakovich MP, Ferruzzi MG, Mitchell CA (2016) Manipulating sensory and phytochemical profiles of greenhouse tomatoes using environmentally relevant doses of ultraviolet radiation. J Agric Food Chem 64(36):6801–6808. https://doi.org/10.1021/acs.jafc.6b02983

Eisenberg T, Abdellatif M, Schroeder S, Primessnig U, Stekovic S, Pendl T, Harger A, Schipke J, Zimmermann A, Schmidt A, Tong M, Ruckenstuhl C, Dammbrueck C, Gross AS, Herbst V, Magnes C, Trausinger G, Narath S, Meinitzer A, Hu Z, Kirsch A, Eller K, Carmona-Gutierrez D, Büttner S, Pietrocola F, Knittelfelder O, Schrepfer E, Rockenfeller P, Simonini C, Rahn A, Horsch M, Moreth K, Beckers J, Fuchs H, Gailus-Durner V, Neff F, Janik D, Rathkolb B, Rozman J, de Angelis MH, Moustafa T, Haemmerle G, Mayr M, Willeit P, von Frieling-Salewsky M, Pieske B, Scorrano L, Pieber T, Pechlaner R, Willeit J, Sigrist SJ, Linke WA, Mühlfeld C, Sadoshima J, Dengjel J, Kiechl S, Kroemer G, Sedej S, Madeo F (2016) Cardioprotection and lifespan extension by the natural polyamine spermidine. Nat Med 22(12):1428–1438. https://doi.org/10.1038/nm.4222

Elbadrawy E, Sello A (2016) Evaluation of nutritional value and antioxidant activity of tomato peel extracts. Arab J Chem 9(2):S1010–S1018

FAO (2019) The state of food security and nutrition in the world. https://www.fao.org/state-of-food-security-nutrition

Fattore M, Montesano D, Pagano E, Teta R, Borrelli F, Mangoni A, Seccia S, Albrizio S (2016) Carotenoid and flavonoid profile and antioxidant activity in "Pomodorino Vesuviano" tomatoes. J Food Compost Anal 53:61–68. https://doi.org/10.1016/j.jfca.2016.08.008

Fawzi W, Herrera MG, Nestel P (2000) Tomato intake in relation to mortality and morbidity among Sudanese children. J Nutr 130(10):2537–2542. https://doi.org/10.1093/jn/130.10.2537

Fei Z, Joung JG, Tang X, Zheng Y, Huang M, Lee JM, McQuinn R, Tieman DM, Alba R, Klee HJ, Giovannoni JJ (2011) Tomato Functional Genomics Database: a comprehensive resource and analysis package for tomato functional genomics. Nucleic Acids Res 39:D1156–D1163. https://doi.org/10.1093/nar/gkq991

Fernández-García E, Carvajal-Lérida I, Jarén-Galán M, Garrido-Fernández J, Pérez-Gálvez A, Hornero-Méndez D (2012) Carotenoids bioavailability from foods: from plant pigments to efficient biological activities. Food Res Int 46(2):438–450. https://doi.org/10.1016/j.foodres.2011.06.007

Fernandez-Pozo N, Menda N, Edwards JD, Saha S, Tecle IY, Strickler SR, Bombarely A, Fisher-York T, Pujar A, Foerster H, Yan A (2015) The Sol Genomics Network (SGN) – from genotype to phenotype to breeding. Nucleic Acids Res 43(D1):D1036–D1041. https://doi.org/10.1093/nar/gku1195

Fernandez-Pozo N, Zheng Y, Snyder SI, Nicolas P, Shinozaki Y, Fei Z, Catala C, Giovannoni JJ, Rose JK, Mueller LA (2017) The tomato expression atlas. Bioinformatics 33(15):2397–2398. https://doi.org/10.1093/bioinformatics/btx190

Fernie AR, Tadmor Y, Zamir D (2006) Natural genetic variation for improving crop quality. Curr Opin Plant Biol 9:196–202. https://doi.org/10.1016/j.pbi.2006.01.010

Figàs MR, Prohens J, Raigón MD, Fita A, García-Martínez MD, Casanova C, Soler S (2015) Characterization of composition traits related to organoleptic and functional quality for the differentiation, selection and enhancement of local varieties of tomato from different cultivar groups. Food Chem 187:517–524. https://doi.org/10.1016/j.foodchem.2015.04.083

Flores P, Sánchez E, Fenoll J, Hellín P (2017) Genotypic variability of carotenoids in traditional tomato cultivars. Food Res Int 100:510–516. https://doi.org/10.1016/j.foodres.2016.07.014

Foerster H, Bombarely A, Battey JN, Sierro N, Ivanov NV, Mueller LA (2018) SolCyc: a database hub at the Sol Genomics Network (SGN) for the manual curation of metabolic networks in *Solanum* and *Nicotiana* specific databases. Database 2018:1–13. https://doi.org/10.1093/database/bay035

Foolad MR (2007) Molecular mapping, marker-assisted selection and map-based cloning in tomato. In: Varshney RK, Tuberosa R (eds) Genomics-assisted crop improvement. Springer, Dordrecht, pp 307–356

Fortes AM, Agudelo-Romero P (2018) Polyamine metabolism in climacteric and non-climacteric fruit ripening. In: Alcázar R, Tiburcio A (eds) Polyamines. Methods in molecular biology, vol 1694. Humana Press, New York, pp 433–447. https://doi.org/10.1007/978-1-4939-7398-9_36

Frary A, Doganlar S, Frampton A, Fulton T, Uhlig J, Yates H, Tanksley S (2003) Fine mapping of quantitative trait loci for improved fruit characteristics from *Lycopersicon chmielewskii* chromosome 1. Genome 46(2):235–243. https://doi.org/10.1139/g02-122

Fraser PD, Kiano JW, Truesdale MR, Schuch W, Bramley PM (1999) Phytoene synthase-2 enzyme activity in tomato does not contribute to carotenoid synthesis in ripening fruit. Plant Mol Biol 40(4):687–698. https://doi.org/10.1023/A:1006256302570

Fraser PD, Romer S, Shipton CA, Mill PB, Kiano JW, Misawa N, Bramley PM (2002) Evaluation of transgenic tomato plants expressing an additional phytoene synthase in a fruit-specific manner. Proc Natl Acad Sci U S A 99(2):1092–1097. https://doi.org/10.1073/pnas.24137459

Friedman M (2002) Tomato glycoalkaloids: role in the plant and in the diet. J Agric Food Chem 50(21):5751–5780. https://doi.org/10.1021/jf020560c

Friedman M (2013) Anticarcinogenic, cardioprotective, and other health benefits of tomato compounds lycopene, α-tomatine, and tomatidine in pure form and in fresh and processed tomatoes. J Agric Food Chem 61(40):9534–9550. https://doi.org/10.1021/jf402654e

Frusciante L, Carli P, Ercolano MR, Pernice R, Di Matteo A, Fogliano V, Pellegrini N (2007) Antioxidant nutritional quality of tomato. Mol Nutr Food Res 51(5):609–617. https://doi.org/10.1002/mnfr.200600158

Gago C, Drosou V, Paschalidis K, Guerreiro A, Miguel G, Antunes D, Hilioti Z (2017) Targeted gene disruption coupled with metabolic screen approach to uncover the LEAFY COTYLEDON1-LIKE4 (L1L4) function in tomato fruit metabolism. Plant Cell Rep 36(7):1065–1082. https://doi.org/10.1007/s00299-017-2137-9

Gale CR, Hall NF, Phillips DIW, Martyn CN (2003) Lutein and zeaxanthin status and risk of age-related macular degeneration. Invest Ophthalmol Vis Sci 44:2461–2465. https://doi.org/10.1167/iovs.02-0929

Gao L, Zhao W, Qu H, Wang Q, Zhao L (2016) The yellow-fruited tomato 1 (yft1) mutant has altered fruit carotenoid accumulation and reduced ethylene production as a result of a genetic lesion in ETHYLENE INSENSITIVE2. Theor Appl Genet 129(4):717–728. https://doi.org/10.1007/s00122-015-2660-4

Gao L, Gonda I, Sun H, Ma Q, Bao K, Tieman DM, Fei Z (2019) The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. Nat Genet 51(6):1044–1051. https://doi.org/10.1038/s41588-019-0410-2

García-Valverde V, Navarro-González I, García-Alonso J, Periago MJ (2013) Antioxidant bioactive compounds in selected industrial processing and fresh consumption tomato cultivars. Food Bioprocess Technol 6(2):391–402. https://doi.org/10.1007/s11947-011-0687-3

Gautier H, Diakou-Verdin V, Bénard C, Reich M, Buret M, Bourgaud F, Poëssel JL, Caris-Veyrat C, Génard M (2008) How does tomato quality (sugar, acid, and nutritional quality) vary with

ripening stage, temperature, and irradiance? J Agric Food Chem 56(4):1241–1250. https://doi.org/10.1021/jf072196t

Gerten D, Heck V, Jägermeyr J, Bodirsky BL, Fetzer I, Jalava M, Kummu M, Lucht W, Rockström J, Schaphoff S, Schellnhuber HJ (2020) Feeding ten billion people is possible within four terrestrial planetary boundaries. Nat Sustain 3:200–208. https://doi.org/10.1038/s41893-019-0465-1

Giampieri F, Battino M (2020) Bioactive phytochemicals and functional food ingredients in fruits and vegetables. Int J Mol Sci 21(9):3278. https://doi.org/10.3390/ijms21093278

Giovinazzo G, D'Amico L, Paradiso A, Bollini R, Sparvoli F, DeGara L (2005) Antioxidant metabolite profiles in tomato fruit constitutively expressing the grapevine stilbene synthase gene. Plant Biotechnol 3(1):57–69. https://doi.org/10.1111/j.1467-7652.2004.00099.x

Giuffrè AM, Capocasale M (2016a) Sterol composition of tomato (*Solanum lycopersicum* L.) seed oil: the effect of cultivar. Int Food Res J 23(1):625

Giuffrè AM, Capocasale M (2016b) Physicochemical composition of tomato seed oil for an edible use: the effect of cultivar. Int Food Res J 23(2):583–591

Go VL, Nguyen CT, Harris DM, Lee WN (2005) Nutrient-gene interaction: metabolic genotype–phenotype relationship. J Nutr 135(12 Suppl):3016S–3020S. https://doi.org/10.1093/jn/135.12.3016s

Goff SA, Klee HJ (2006) Plant volatile compounds: sensory cues for health and nutritional value? Science 311:815–819. https://doi.org/10.1126/science.1112614

Gouranton E, Thabuis C, Riollet C, Malezet-Desmoulins C, El Yazidi C, Amiot MJ, Borel P, Landrier JF (2011) Lycopene inhibits proinflammatory cytokine and chemokine expression in adipose tissue. J Nutr Biochem 22(7):642–648. https://doi.org/10.1016/j.jnutbio.2010.04.016

Gramazio P, Takayama M, Ezura H (2020) Challenges and prospects of new plant breeding techniques for GABA improvement in crops: tomato as an example. Front Plant Sci 11:577980. https://doi.org/10.3389/fpls.2020.577980

Grandillo S, Chetelat R, Knapp S, Spooner D, Peralta I, Cammareri M, Perez O, Termolino P, Tripodi P, Chiusano ML, Ercolano MR (2011) *Solanum* sect. *Lycopersicon*. In: Kole C (ed) Wild crop relatives: genomic and breeding resources. Springer, New York, pp 129–215

Grassino AN, Halambek J, Djakovic S, Brncic SR, Dent M, Grabaric Z (2016) Utilization of tomato peel waste from canning factory as a potential source for pectin production and application as tin corrosion inhibitor. Food Hydrocoll 52:265–274. https://doi.org/10.1016/j.foodhyd.2015.06.020

Guo F, Zhou W, Zhang J, Xu DX (2012) Effect of the citrus lycopene β-cyclase transgene on carotenoid metabolism in transgenic tomato fruits. PLoS One 7(2):e32221. https://doi.org/10.1371/journal.pone.0032221

Gutierrez EE, Johansson E, Centellas Quezada A, Gustavsson KE, Olsson ME (2021) Genotype and maturity stage affect the content and composition of polyamines in tomato – possible relations to plant and human health. Horticulturae 7(9):300. https://doi.org/10.3390/horticulturae7090300

Hall RD, Brouwer ID, Fitzgerald MA (2008) Plant metabolomics and its potential application for human nutrition. Physiol Plant 132(2):162–175. https://doi.org/10.1111/j.1399-3054.2007.00989.x

Hanif R, Iqbal Z, Iqbal M, Hanif S, Rasheed M (2006) Use of vegetables as nutritional food: role in human health. J Agric Biol Sci 1(1):18–22

Hano S, Shibuya T, Imoto N, Ito A, Imanishi S, Aso H et al (2017) Serotonin content in fresh and processed tomatoes and its accumulation during fruit development. Sci Hortic 214:107–113. https://doi.org/10.1016/j.scienta.2016.11.009

Hanson PM, Yang RY, Wu J, Chen JT, Ledesma D, Tsou SC, Lee TC (2004) Variation for antioxidant activity and antioxidants in tomato. J Am Soc Hortic Sci 129(5):704–711. https://doi.org/10.21273/JASHS.129.5.0704

Haroldsen VM, Chi-Ham CL, Kulkarni S, Lorence A, Bennett AB (2011) Constitutively expressed DHAR and MDHAR influence fruit, but not foliar ascorbate levels in tomato. Plant Physiol Biochem 49(10):1244–1249. https://doi.org/10.1016/j.plaphy.2011.08.003

Hartwell LH, Goldberg ML, Fischer JA, Hood L (2018) Genetics from genes to genomes, 6th edn. McGraw-Hill, New York

Hasegawa S, Jao TM, Inagi R (2017) Dietary metabolites and chronic kidney disease. Nutrients 9(4):358. https://doi.org/10.3390/nu9040358

Hawkins C, Ginzburg D, Zhao K, Dwyer W, Xue B, Xu A, Rice S, Cole B, Paley S, Karp P, Rhee SY (2021) Plant Metabolic Network 15: a resource of genome-wide metabolism databases for 126 plants and algae. J Integr Plant Biol 63(11):1888–1905. https://doi.org/10.1111/jipb.13163

Herrera PG, Sánchez-Mata MC, Cámara M (2010) Nutritional characterization of tomato fiber as a useful ingredient for food industry. Innov Food Sci Emerg Technol 11(4):707–711. https://doi.org/10.1016/j.ifset.2010.07.005

Hilioti Z, Ganopoulos I, Ajith S, Bossis I, Tsaftaris A (2016) A novel arrangement of zinc finger nuclease system for in vivo targeted genome engineering: the tomato LEC1-LIKE4 gene case. Plant Cell Rep 35(11):2241–2255. https://doi.org/10.1007/s00299-016-2031-x

Holzapfel NP, Holzapfel BM, Champ S, Feldthusen J, Clements J, Hutmacher DW (2013) The potential role of lycopene for the prevention and therapy of prostate cancer: from molecular mechanisms to clinical evidence. Int J Mol Sci 14(7):14620–14646. https://doi.org/10.3390/ijms140714620

Huang S, Weigel D, Beachy RN, Li J (2016) A proposed regulatory framework for genome-edited crops. Nat Genet 48(2):109–111. https://doi.org/10.1038/ng.3484

Huché-Thélier L, Crespel L, Le Gourrierec J, Morel P, Sakr S, Leduc N (2016) Light signaling and plant responses to blue and UV radiations: perspectives for applications in horticulture. Environ Exp Bot 121:22–38. https://doi.org/10.1016/j.envexpbot.2015.06.009

Huh MK, Youn SJ, Kang SC (2011) Identification and genetic diversity of Korean tomato cultivars by RAPD markers. J Life Sci 21(1):15–21. https://doi.org/10.5352/JLS.2011.21.1.15

Iijima Y, Watanabe B, Sasaki R, Takenaka M, Ono H, Sakurai N, Umemoto N, Suzuki H, Shibata D, Aoki K (2013) Steroidal glycoalkaloid profiling and structures of glycoalkaloids in wild tomato fruit. Phytochemistry 95:145–157. https://doi.org/10.1016/j.phytochem.2013.07.016

Ilahy R, Piro G, Tlili I, Riahi A, Sihem R, Ouerghi I, Hdider C, Lenucci MS (2016) Fractionate analysis of the phytochemical composition and antioxidant activities in advanced breeding lines of high-lycopene tomatoes. Food Funct 7(1):574–583. https://doi.org/10.1039/C5FO00553A

Ilahy R, Siddiqui MW, Piro G, Lenucci MS, Hdider C, Helyes L (2017) A focus on high-lycopene tomato cultivars: horticultural performance and functional quality. Acta Hortic 1159:57–64. https://doi.org/10.17660/actahortic.2017.1159

Iniesta MD, Pérez-Conesa D, García-Alonso J, Ros G, Periago MJ (2009) Folate content in tomato (*Lycopersicon esculentum*). Influence of cultivar, ripeness, year of harvest, and pasteurization and storage temperatures. J Agric Food Chem 57(11):4739–4745. https://doi.org/10.1021/jf900363r

Isaacson T, Ronen G, Zamir D, Hirschberg J (2002) Cloning of tangerine from tomato reveals a carotenoid isomerase essential for the production of β-carotene and xanthophylls in plants. Plant Cell 14(2):333–342. https://doi.org/10.1105/tpc.010303

Ishitani M, Rao I, Wenzl P, Beebe S, Tohme J (2004) Integration of genomics approach with traditional breeding towards improving abiotic stress adaptation: drought and aluminum toxicity as case studies. Field Crop Res 90:35–45. https://doi.org/10.1016/j.fcr.2004.07.004

Itkin M, Heinig U, Tzfadia O, Bhide AJ, Shinde B, Cardenas PD, Bocobza SE, Unger T, Malitsky S, Finkers R, Tikunov Y (2013) Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. Science 341(6142):175–179. https://doi.org/10.1126/science.1240230

Jian W, Cao H, Yuan S, Liu Y, Lu J, Lu W, Li Z (2019) SlMYB75, an MYB-type transcription factor, promotes anthocyanin accumulation and enhances volatile aroma production in tomato fruits. Hortic Res 6. https://doi.org/10.1038/s41438-018-0098-y

Kabelka E, Yang W, Francis DM (2004) Improved tomato fruit color within an inbred backcross line derived from *Lycopersicon esculentum* and *L. hirsutum* involves the interaction of loci. J Am Soc Hortic Sci 129(2):250–257. https://doi.org/10.21273/JASHS.129.2.0250

Karppi J, Kurl S, Mäkikallio TH, Ronkainen K, Laukkanen JA (2013) Serum β-carotene concentrations and the risk of congestive heart failure in men: a population-based study. Int J Cardiol 168(3):1841–1846. https://doi.org/10.1016/j.ijcard.2012.12.072

Kavitha P, Shivashankara KS, Rao VK, Sadashiva AT, Ravishankar KV, Sathish GJ (2014) Genotypic variability for antioxidant and quality parameters among tomato cultivars, hybrids, cherry tomatoes and wild species. J Sci Food Agric 94(5):993–999. https://doi.org/10.1002/jsfa.6359

Kazachkova Y, Zemach I, Panda S, Bocobza S, Vainer A, Rogachev I, Dong Y, Ben-Dor S, Veres D, Kanstrup C, Lambertz SK, Aharoni A (2021) The GORKY glycoalkaloid transporter is indispensable for preventing tomato bitterness. Nat Plants 7(4):468–480. https://doi.org/10.1038/s41477-021-00865-6

Kim JE, Oh SK, Lee JH, Lee BM, Jo SH (2014) Genome-wide SNP calling using next generation sequencing data in tomato. Mol Cells 37(1):36–42. https://doi.org/10.14348/molcells.2014.2241

Kim M, Nguyen TTP, Ahn JH, Kim GJ, Sim SC (2021) Genome-wide association study identifies QTL for eight fruit traits in cultivated tomato (*Solanum lycopersicum* L.). Hortic Res 8(1):203. https://doi.org/10.1038/s41438-021-00638-4

Knapp S, Peralta IE (2016) The tomato (*Solanum lycopersicum* L., *Solanaceae*) and its botanical relatives. In: Causse M, Giovannoni J, Bouzayen M, Zouine M (eds) The tomato genome. Springer, Berlin, pp 7–21

Koh E, Charoenprasert S, Mitchell AE (2012) Effects of industrial tomato paste processing on ascorbic acid, flavonoids and carotenoids and their stability over one-year storage. J Sci Food Agric 92(1):23–28. https://doi.org/10.1002/jsfa.4580

Koike S, Matsukura C, Takayama M, Asamizu E, Ezura H (2013) Suppression of γ-aminobutyric acid (GABA) transaminases induces prominent GABA accumulation, dwarfism and infertility in the tomato (*Solanum lycopersicum* L.). Plant Cell Physiol 54(5):793–807. https://doi.org/10.1093/pcp/pct035

Korir NK, Diao W, Tao R, Li X, Kayesh E, Li A, Wang S (2014) Genetic diversity and relationships among different tomato varieties revealed by EST-SSR markers. Genet Mol Res 13(1):43–53. https://doi.org/10.4238/2014.January.8.3

Kovacs K, Zhang L, Linforth RS, Whittaker B, Hayes CJ, Fray RG (2007) Redirection of carotenoid metabolism for the efficient production of taxadiene [taxa-4 (5), 11 (12)-diene] in transgenic tomato fruit. Transgenic Res 16(1):121–126. https://doi.org/10.1007/s11248-006-9039-x

Kozukue N, Han JS, Lee KR, Friedman M (2004) Dehydrotomatine and α-tomatine content in tomato fruits and vegetative plant tissues. J Agric Food Chem 52(7):2079–2083. https://doi.org/10.1021/jf0306845

Kudo T, Kobayashi M, Terashima S, Katayama M, Ozaki S, Kanno M, Saito M, Yokoyama K, Ohyanagi H, Aoki K, Kubo Y (2017) TOMATOMICS: a web database for integrated omics information in tomato. Plant Cell Physiol 58(1):e8. https://doi.org/10.1093/pcp/pcw207

Kusano M, Fukushima A (2013) Current challenges and future potential of tomato breeding using omics approaches. Breed Sci 63(1):31–41. https://doi.org/10.1270/jsbbs.63.31

Kuwabara K, Harada I, Matsuzawa Y, Ariizumi T, Shirasawa K (2021) Organelle genome assembly uncovers the dynamic genome reorganization and cytoplasmic male sterility associated genes in tomato. Hortic Res 8:250. https://doi.org/10.1038/s41438-021-00676-y

Lattanzio V (2013) Phenolic compounds: introduction. In: Ramawat KG, Merillon JM (eds) Natural products. Springer, Berlin, pp 1543–1580. https://doi.org/10.1007/978-3-642-22144-6_57

Lecomte L, Duffe P, Buret M, Servin B, Causse M (2004) Marker-assisted introgression of five QTLs controlling fruit quality traits into three tomato lines revealed interactions between QTLs and genetic backgrounds. Theor Appl Genet 109(3):658–668. https://doi.org/10.1007/s00122-004-1674-0

Lenucci MS, Cadinu D, Taurino M, Piro G, Dalessandro G (2006) Antioxidant composition in cherry and high-pigment tomato cultivars. J Agric Food Chem 54(7):2606–2613. https://doi.org/10.1021/jf052920c

Leonardi C, Ambrosino P, Esposito F, Fogliano V (2000) Antioxidative activity and carotenoid and tomatine contents in different typologies of fresh consumption tomatoes. J Agric Food Chem 48(10):4723–4727. https://doi.org/10.1021/jf000225t

Levin I, de Vos CHR, Tadmor Y, Bovy A, Lieberman M, Oren-Shamir M, Segev O, Kolotilin I, Keller M, Ovadia R, Meir A, Bino RJ (2003) High pigment tomato mutants – more than just lycopene. Isr J Plant Sci 54:179–190. https://doi.org/10.1560/IJPS_54_3_179

Li Q, Li Y, Li C, Yu X (2012) Enhanced ascorbic acid accumulation through overexpression of dehydroascorbate reductase confers tolerance to methyl viologen and salt stresses in tomato. Czech J Genet Plant Breed 48(2):74–86

Li R, Li R, Li X, Fu D, Zhu B, Tian H, Luo Y, Zhu H (2018a) Multiplexed CRISPR/Cas9-mediated metabolic engineering of γ-aminobutyric acid levels in *Solanum lycopersicum*. Plant Biotechnol J 16(2):415–427. https://doi.org/10.1111/pbi.12781

Li X, Wang Y, Chen S, Tian H, Fu D, Zhu B, Zhu H (2018b) Lycopene is enriched in tomato fruit by CRISPR/Cas9-mediated multiplex genome editing. Front Plant Sci 9:559. https://doi.org/10.3389/fpls.2018.00559

Li N, Wu H, Ding Q, Li H, Li Z, Ding J, Li Y (2018c) The heterologous expression of *Arabidopsis* PAP2 induces anthocyanin accumulation and inhibits plant growth in tomato. Funct Integr Genomics 18(3):341–353. https://doi.org/10.1007/s10142-018-0590-3

Li G, Liu YG, Chen Y (2019) Genome-editing technologies: the gap between application and policy. Sci China Life Sci 62(11):1534–1538. https://doi.org/10.1007/s11427-019-1566-1

Li J, Scarano A, Gonzalez NM, D'Orso F, Yue Y, Nemeth K, Saalbach G, Hill L, de Oliveira Martins C, Moran R, Santino A (2022) Biofortified tomatoes provide a new route to vitamin D sufficiency. Nat Plants 8:611–616. https://doi.org/10.1038/s41477-022-01154-6

Lim MY, Pulla RK, Park JM, Harn CH, Jeong BR (2012) Over-expression of L-gulono-γ-lactone oxidase (GLOase) gene leads to ascorbate accumulation with enhanced abiotic stress tolerance in tomato. In Vitro Cell Dev Biol Plant 48(5):453–461. https://doi.org/10.1007/s11627-012-9461-0

Lin D, Xiao M, Zhao J, Li Z, Xing B, Li X, Kong M, Li L, Zhang Q, Liu Y, Chen H (2016) An overview of plant phenolic compounds and their importance in human nutrition and management of type 2 diabetes. Molecules 21(10):1374. https://doi.org/10.3390/molecules21101374

Lincoln RE, Zscheile FP, Porter JW, Kohler GW, Caldwell RM (1943) Provitamin A and vitamin C in the genus *Lycopersicon*. Bot Gaz 105(1):113–115

Liu L, Jia C, Zhang M, Chen D, Chen S, Guo R, Wang Q (2014) Ectopic expression of a BZR1-1D transcription factor in brassinosteroid signalling enhances carotenoid accumulation and fruit quality attributes in tomato. Plant Biotechnol J 12(1):105–115. https://doi.org/10.1111/pbi.12121

Loiudice R, Impembo M, Laratta B, Villari G, Voi AL, Siviero P, Castaldo D (1995) Composition of San Marzano tomato varieties. Food Chem 53(1):81–89. https://doi.org/10.1016/0308-8146(95)95791-4

Lu Z, Wang J, Gao R, Ye F, Zhao G (2019) Sustainable valorisation of tomato pomace: a comprehensive review. Trends Food Sci Technol 86:172–187. https://doi.org/10.1016/j.tifs.2019.02.020

Luthria DL, Mukhopadhyay S, Krizek DT (2006) Content of total phenolics and phenolic acids in tomato (*Lycopersicon esculentum* Mill.) fruits as influenced by cultivar and solar UV radiation. J Food Compos Anal 19(8):771–777. https://doi.org/10.1016/j.jfca.2006.04.005

Madeo F, Eisenberg T, Pietrocola F, Kroemer G (2018) Spermidine in health and disease. Science 359(6374). https://doi.org/10.1126/science.aan2788

Madhulatha P, Gupta A, Gupta S, Kumar A, Pal RK, Rajam MV (2014) Fruit-specific overexpression of human S-adenosylmethionine decarboxylase gene results in polyamine accumulation and affects diverse aspects of tomato fruit development and quality. J Plant Biochem Biotechnol 23(2):151–160. https://doi.org/10.1007/s13562-013-0194-x

Mahjoub A, Hernould M, Joubès J, Decendit A, Mars M, Barrieu F, Delrot S (2009) Overexpression of a grapevine R2R3-MYB factor in tomato affects vegetative development, flower morphology and flavonoid and terpenoid metabolism. Plant Physiol Biochem 47(7):551–561. https://doi.org/10.1016/j.plaphy.2009.02.015

Manganaris GA, Goulas V, Mellidou I, Drogoudi P (2018) Antioxidant phytochemicals in fresh produce: exploitation of genotype variation and advancements in analytical protocols. Front Chem 6(5):95. https://doi.org/10.3389/fchem.2017.00095

Manoharan RK, Jung HJ, Hwang I, Jeong N, Kho KH, Chung MY, Nou IS (2017) Molecular breeding of a novel orange-brown tomato fruit with enhanced beta-carotene and chlorophyll accumulation. Hereditas 154:1. https://doi.org/10.1186/s41065-016-0023-z

Mao Y, Botella JR, Liu Y, Zhu JK (2019) Gene editing in plants: progress and challenges. Natl Sci Rev 6(3):421–437. https://doi.org/10.1093/nsr/nwz005

Marsić KN, Sircelj H, Kastelec D (2010) Lipophilic antioxidants and some carpometric characteristics of fruits of ten processing tomato varieties, grown in different climatic conditions. J Agric Food Chem 58(1):390–397. https://doi.org/10.1021/jf902465e

Martí R, Roselló S, Cebolla-Cornejo J (2016) Tomato as a source of carotenoids and polyphenols targeted to cancer prevention. Cancers 8(6):58. https://doi.org/10.3390/cancers8060058

Martí R, Leiva-Brondo M, Lahoz I, Campillo C, Cebolla-Cornejo J, Roselló S (2018) Polyphenol and L-ascorbic acid content in tomato as influenced by high lycopene genotypes and organic farming at different environments. Food Chem 239:148–156. https://doi.org/10.1016/j.foodchem.2017.06.102

Martínez-Valverde I, Periago MJ, Provan G, Chesson A (2002) Phenolic compounds, lycopene and antioxidant activity in commercial varieties of tomato (*Lycopersicum esculentum*). J Sci Food Agric 82(3):323–330. https://doi.org/10.1002/jsfa.1035

Mattoo AK, Chung SH, Goyal RK, Fatima T, Solomos T, Srivastava A, Handa AK (2007) Overaccumulation of higher polyamines in ripening transgenic tomato fruit revives metabolic memory, upregulates anabolism-related genes, and positively impacts nutritional quality. J AOAC Int 90(5):1456–1464. https://doi.org/10.1093/jaoac/90.5.1456

McGloughlin MN (2010) Modifying agricultural crops for improved nutrition. N Biotechnol 27(5):494–504. https://doi.org/10.1016/j.nbt.2010.07.013

Mehta RA, Cassol T, Li N, Ali N, Handa AK, Mattoo AK (2002) Engineered polyamine accumulation in tomato enhances phytonutrient content, juice quality, and vine life. Nat Biotechnol 20:613–618. https://doi.org/10.1038/nbt0602-613

Mellidou I, Koukounaras A, Kostas S, Patelou E, Kanellis AK (2021) Regulation of vitamin C accumulation for improved tomato fruit quality and alleviation of abiotic stress. Genes 12(5):694. https://doi.org/10.3390/genes12050694

Mène-Saffrané L (2017) Vitamin E biosynthesis and its regulation in plants. Antioxidants 7(1):2. https://doi.org/10.3390/antiox7010002

Merenkova S, Zinina O, Stuart M, Okuskhanova E, Androsova N (2020) Effects of human health: a review. Hum Sport Med 20:106–113

Mi J, Vallarino JG, Petřík I, Novák O, Correa SM, Chodasiewicz M, Moreno JC (2022) A manipulation of carotenoid metabolism influence biomass partitioning and fitness in tomato. Metab Eng 70:166–180. https://doi.org/10.1016/j.ymben.2022.01.004

Miller AP, Coronel J, Amengual J (2020) The role of β-carotene and vitamin A in atherogenesis: evidences from preclinical and clinical studies. Biochim Biophys Acta Mol Cell Biol Lipids 1865(11):158635. https://doi.org/10.1016/j.bbalip.2020.158635

Minoggio M, Bramati L, Simonetti P, Gardana C, Iemoli L, Santangelo E, Pietta PG (2003) Polyphenol pattern and antioxidant activity of different tomato lines and cultivars. Ann Nutr Metab 47(2):64–69. https://doi.org/10.1159/000069277

Moco S, Bino RJ, Vorst O, Verhoeven HA, de Groot J, van Beek TA, Vervoort J, De Vos CR (2006) A liquid chromatography-mass spectrometry-based metabolome database for tomato. Plant Physiol 141(4):1205–1218. https://doi.org/10.1104/pp.106.078428

Monforte A, Tanksley S (2000) Fine mapping of a quantitative trait locus (QTL) from *Lycopersicon hirsutum* chromosome 1 affecting fruit characteristics and agronomic traits: breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. Theor Appl Genet 100:471–479. https://doi.org/10.1007/s001220050061

Montesinos-Pereira D, Barrameda-Medina Y, Romero L, Ruiz JM, Sánchez-Rodríguez E (2014) Genotype differences in the metabolism of proline and polyamines under moderate drought in tomato plants. Plant Biol 16(6):1050–1057. https://doi.org/10.1111/plb.12178

Montonen J, Knekt P, Järvinen R, Reunanen A (2004) Dietary antioxidant intake and risk of type 2 diabetes. Diabetes Care 27(2):362–366. https://doi.org/10.2337/diacare.27.2.362

Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiol 47:969–977. https://doi.org/10.1104/pp.108.118232

Morilla A, García JM, Albi MA (1996) Free polyamine contents and decarboxylase activities during tomato development and ripening. J Agric Food Chem 44(9):2608–2611. https://doi.org/10.1021/jf950672m

Mou W, Li D, Luo Z, Mao L, Ying T (2015) Transcriptomic analysis reveals possible influences of ABA on secondary metabolism of pigments, flavonoids and antioxidants in tomato fruit during ripening. PLoS One 10(6):e0129598. https://doi.org/10.1371/journal.pone.0129598

Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y, Herbst EV (2005) The SOL Genomics Network. A comparative resource for *Solanaceae* biology and beyond. Plant Physiol 138(3):310–1317. https://doi.org/10.1104/pp.105.060707

Muñoz-Esparza NC, Latorre-Moratalla ML, Comas-Basté O, Toro-Funes N, Veciana-Nogués MT, Vidal-Carou MC (2019) Polyamines in food. Front Nutr 6:108. https://doi.org/10.3389/fnut.2019.00108

Nakayasu M, Shioya N, Shikata M, Thagun C, Abdelkareem A, Okabe Y, Ariizumi T, Arimura GI, Mizutani M, Ezura H, Hashimoto T (2018) JRE 4 is a master transcriptional regulator of defense-related steroidal glycoalkaloids in tomato. Plant J 94(6):975–990. https://doi.org/10.1111/tpj.13911

Nakayasu M, Akiyama R, Kobayashi M, Lee HJ, Kawasaki T, Watanabe B, Urakawa S, Kato J, Sugimoto Y, Iijima Y, Saito K (2020) Identification of α-tomatine 23-hydroxylase involved in the detoxification of a bitter glycoalkaloid. Plant Cell Physiol 61(1):21–28. https://doi.org/10.1093/pcp/pcz224

Nambeesan S, Datsenka T, Ferruzzi MG, Malladi A, Mattoo AK, Handa AK (2010) Overexpression of yeast spermidine synthase impacts ripening, senescence and decay symptoms in tomato. Plant J 63(5):836–847. https://doi.org/10.1111/j.1365-313X.2010.04286.x

Narise T, Sakurai N, Obayashi T, Ohta H, Shibata D (2017) Co-expressed Pathways DataBase for Tomato: a database to predict pathways relevant to a query gene. BMC Genomics 18(1):437. https://doi.org/10.1186/s12864-017-3786-3

Natalini A, Acciarri N, Cardi T (2021) Breeding for nutritional and organoleptic quality in vegetable crops: the case of tomato and cauliflower. Agriculture 11(7):606. https://doi.org/10.3390/agriculture11070606

Neily MH, Matsukura C, Maucourt M, Bernillon S, Deborde C, Moing A, Yin YG, Saito T, Mori K, Asamizu E, Rolin D (2011) Enhanced polyamine accumulation alters carotenoid metabolism at the transcriptional level in tomato fruit over-expressing spermidine synthase. J Plant Physiol 168(3):242–252. https://doi.org/10.1016/j.jplph.2010.07.003

Nonaka S, Arai C, Takayama M, Matsukura C, Ezura H (2017) Efficient increase of ɣ-aminobutyric acid (GABA) content in tomato fruits by targeted mutagenesis. Sci Rep 7:7057. https://doi.org/10.1038/s41598-017-06400-y

Nour V, Trandafir I, Ionica ME (2014) Evolution of antioxidant activity and bioactive compounds in tomato (*Lycopersicon esculentum* Mill.) fruits during growth and ripening. J Appl Bot Food Qual 87:97–103. https://doi.org/10.5073/JABFQ.2014.087.015

Ökmen B, Şiğva HÖ, Gürbüz N, Ülger M, Frary A, Doğanlar S (2011) Quantitative trait loci (QTL) analysis for antioxidant and agronomically important traits in tomato (*Lycopersicon esculentum*). Turk J Agric For 35(5):501–514. https://doi.org/10.3906/tar-1008-1207

Olaniyi JO, Akanbi WB, Adejumo TA, Ak OG (2010) Growth, fruit yield and nutritional quality of tomato varieties. Afr J Food Sci 4(6):398–402

Opara UL, Al-Ani MR, Al-Rahbi NM (2012) Effect of fruit ripening stage on physico-chemical properties, nutritional composition and antioxidant components of tomato (*Lycopersicum esculentum*) cultivars. Food Bioprocess Technol 5:3236–3243. https://doi.org/10.1007/s11947-011-0693-5

Ortega-Ortiz H, Benavides-Mendoza A, Mendoza-Villarreal R, Ramírezrodríguez H, Romenus KDA (2007) Enzymatic activity in tomato fruits as a response to chemical elicitors. J Mex Chem Soc 51:141–144. https://doi.org/10.29356/JMCS.V51I3.1343

Orzaez D, Medina A, Torre S, Fernández-Moreno JP, Rambla JL, Fernández-del-Carmen A, Granell A (2009) A visual reporter system for virus-induced gene silencing in tomato fruit based on anthocyanin accumulation. Plant Physiol 150(3):1122–1134. https://doi.org/10.1104/pp.109.139006

Pandey A, Misra P, Choudhary D, Yadav R, Goel R, Bhambhani S, Kumar Trivedi P (2015) AtMYB12 expression in tomato leads to large scale differential modulation in transcriptome and flavonoid content in leaf and fruit tissues. Sci Rep 5:12412. https://doi.org/10.1038/srep12412

Park YH, West MA, St Clair DA (2004) Evaluation of AFLPs for germplasm fingerprinting and assessment of genetic diversity in cultivars of tomato (*Lycopersicon esculentum* L.). Genome 47(3):510–518. https://doi.org/10.1139/g04-00

Pék Z, Szuvandzsiev P, Daood H, Neményi A, Helyes L (2014) Effect of irrigation on yield parameters and antioxidant profiles of processing cherry tomato. Cent Eur J Biol 9:383–395. https://doi.org/10.2478/s11535-013-0279-5

Peralta IE, Spooner DM (2007) History, origin and early cultivation of tomato (*Solanaceae*). In: Razdan MK, Mattoo AK (eds) Genetic improvement of solanaceous crops, vol 2. Tomato. Science Publishers, Enfield, pp 1–27

Peralta IE, Knapp S, Spooner DM (2005) New species of wild tomatoes (*Solanum* section *Lycopersicon*: *Solanaceae*) from northern Peru. Syst Bot 30(2):424–434. https://doi.org/10.1600/0363644054223657

Peralta IE, Spooner DM, Knapp S (2008) Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicoides*, sect. *Juglandifolia*, sect. *Lycopersicon*; *Solanaceae*). Monogr Syst Bot 84. https://doi.org/10.2307/25027972

Perez-Fons L, Wells T, Corol DI, Ward JL, Gerrish C, Beale MH, Fraser PD (2014) A genome-wide metabolomic resource for tomato fruit from *Solanum pennellii*. Sci Rep 4:3859. https://doi.org/10.1038/srep03859

Phan NT, Sim SC (2017) Genomic tools and their implications for vegetable breeding. Hortic Sci Technol 35(2):149–164. https://doi.org/10.12972/kjhst.20170018

Picó B, Herráiz J, Nuez F (2000) *L. chilense*-derived bridge lines for introgressing peruvianum traits in esculentum genome. Rep Tomato Genet Coop 50:30–32

Raiola A, Rigano MM, Calafiore R, Frusciante L, Barone A (2014) Enhancing the health-promoting effects of tomato fruit for biofortified food. Mediat Inflamm 2014:139873. https://doi.org/10.1155/2014/139873

Raiola A, Tenore GC, Barone A, Frusciante L, Rigano MM (2015) Vitamin E content and composition in tomato fruits: Beneficial roles and bio-fortification. Int J Mol Sci 16(112):29250–29264. https://doi.org/10.3390/ijms161226163

Ramos-Bueno RP, Romero-González R, González-Fernández MJ, Guil-Guerrero JL (2017) Phytochemical composition and in vitro anti-tumour activities of selected tomato varieties. J Sci Food Agric 97(2):488–496. https://doi.org/10.1002/jsfa.7750

Rao BN (2003) Bioactive phytochemicals in Indian foods and their potential in health promotion and disease prevention. Asia Pac J Clin Nutr 12(1):9–22

Rao AV, Rao LG (2007) Carotenoids and human health. Pharmacol Res 55(3):207–216. https://doi.org/10.1016/j.phrs.2007.01.012

Rapa M, Ciano S, Ruggieri R, Vinci G (2021) Bioactive compounds in cherry tomatoes (*Solanum lycopersicum* var. *cerasiforme*): cultivation techniques classification by multivariate analysis. Food Chem 355:129630. https://doi.org/10.1016/j.foodchem.2021.129630

Rick CM (1951) Hybrids between *Lycopersicon esculentum* Mill. and *Solanum lycopersicoides* Dun. Proc Natl Acad Sci U S A 37:741–744. https://doi.org/10.1073/pnas.37.11.741

Rick CM (1973) Potential genetic resources in tomato species: clues from observations in native habitats. In: Srb AM (ed) Genes, enzymes and populations. Plenum Press, New York, pp 255–269

Rick CM (1980) Tomato. In: Fehr WR, Hadley HH (eds) Hybridization of crop plants. American Society of Agronomy, Madison, pp 669–680

Rick CM (1983) Crossability between *L. esculentum* and a new race of *L. peruvianum*. Rep Tomato Genet Coop 33:13

Rick CM, Chetelat RT (1995) Utilization of related wild species for tomato improvement. Int Symp Solanacea Fresh Mark 412:21–38. https://doi.org/10.17660/ActaHortic.1995.412.1

Rigano MM, Raiola A, Tenore GC, Monti DM, Del Giudice R, Frusciante L, Barone A (2014) Quantitative trait loci pyramiding can improve the nutritional potential of tomato (*Solanum lycopersicum*) fruits. J Agric Food Chem 62(47):11519–11527. https://doi.org/10.1021/jf502573n

Ronen G, Cohen M, Zamir D, Hirschberg J (1999) Regulation of carotenoid biosynthesis during tomato fruit development: expression of the gene for lycopene epsilon-cyclase is down-regulated during ripening and is elevated in the mutant Delta. Plant J 17(4):341–351. https://doi.org/10.1046/j.1365-313X.1999.00381.x

Ronen G, Carmel-Goren L, Zamir D, Hirschberg J (2000) An alternative pathway to β-carotene formation in plant chromoplasts discovered by map-based cloning of Beta and old-gold color mutations in tomato. Proc Natl Acad Sci U S A 97(20):11102–11107. https://doi.org/10.1073/pnas.19017749

Rosati C, Aquilani R, Dharmapuri S, Pallara P, Marusic C, Tavazza R, Giuliano G (2000) Metabolic engineering of beta-carotene and lycopene content in tomato fruit. Plant J 24(3):413–420. https://doi.org/10.1046/j.1365-313x.2000.00880.x

Roselló S, Adalid AM, Cebolla-Cornejo J, Nuez F (2011) Evaluation of the genotype, environment and their interaction on carotenoid and ascorbic acid accumulation in tomato germplasm. J Sci Food Agric 91(6):1014–1021. https://doi.org/10.1002/jsfa.4276

Rousseaux MC, Jones CM, Adams D, Chetelat R, Bennett A, Powell A (2005) QTL analysis of fruit antioxidants in tomato using *Lycopersicon pennellii* introgression lines. Theor Appl Genet 111: 1396–1408. https://doi.org/10.1007/s00122-005-0071-7

Ruggieri V, Francese G, Sacco A, D'Alessandro A, Rigano MM, Parisi M, Barone A (2014) An association mapping approach to identify favourable alleles for tomato fruit quality breeding. BMC Plant Biol 14:337. https://doi.org/10.1186/s12870-014-0337-9

Ruggieri V, Bostan H, Barone A, Frusciante L, Chiusano ML (2016) Integrated bioinformatics to decipher the ascorbic acid metabolic network in tomato. Plant Mol Biol 91:397–412. https://doi.org/10.1007/s11103-016-0469-4

Sacco A, Di Matteo A, Lombardi N, Trotta N, Punzo B, Mari A, Barone A (2013) Quantitative trait loci pyramiding for fruit quality traits in tomato. Mol Breed 31:217–222. https://doi.org/10.1007/s11032-012-9763-2

Sacco A, Raiola A, Calafiore R, Barone A, Rigano MM (2019) New insights in the control of antioxidants accumulation in tomato by transcriptomic analyses of genotypes exhibiting contrasting levels of fruit metabolites. BMC Genomics 20:43. https://doi.org/10.1186/s12864-019-5428-4

Saito T, Matsukura C, Sugiyama M, Watahiki A, Ohshima I, Iijima Y, Konishi C, Fujii T, Inai S, Fukuda N, Nishimura S (2008) Screening for γ-aminobutyric acid (GABA)-rich tomato varieties. J Jpn Soc Hortic Sci 77(3):242–250. https://doi.org/10.2503/jjshs1.77.242

Salava H, Thula S, Mohan V, Kumar R, Maghuly F (2021) Application of genome editing in tomato breeding: mechanisms, advances, and prospects. Int J Mol Sci 22(2):682. https://doi.org/10.3390/ijms22020682

Salunke DS, Jadhav AS, Pawar BD, Kale AA, Chimote VP (2012) Diversity analysis of tomato genotypes using RAPD markers and high-performance liquid chromatography in relation to β-carotene content. Vegetos 25(2):95–101

SanGiovanni JP, Chew EY, Clemons TE, Ferris FL III, Gensler G, Lindblad AS, Milton RC, Seddon JM, Sperduto RD (2007) The relationship of dietary carotenoid and vitamin A, E, and C intake with age-related macular degeneration in a case-control study: AREDS report no. 22. Arch Ophthalmol 125(9):1225–1232. https://doi.org/10.1001/archopht.125.9.1225

Schauer N, Semel Y, Balbo I, Steinfath M, Repsilber D, Selbig J, Fernie AR (2008) Mode of inheritance of primary metabolic traits in tomato. Plant Cell 20(3):509–523. https://doi.org/10.1105/tpc.107.056523

Schijlen E, Ric de Vos CH, Jonker H, Van Den Broeck H, Molthoff J, Van Tunen A, Bovy A (2006) Pathway engineering for healthy phytochemicals leading to the production of novel flavonoids in tomato fruit. Plant Biotechnol J 4(4):433–444. https://doi.org/10.1111/j.1467-7652.2006.00192.x

Schreiber G, Reuveni M, Evenor D, Oren-Shamir M, Ovadia R, Sapir-Mir M, Levin I (2012) ANTHOCYANIN1 from *Solanum chilense* is more efficient in accumulating anthocyanin metabolites than its *Solanum lycopersicum* counterpart in association with the ANTHOCYANIN FRUIT phenotype of tomato. Theor Appl Genet 124:295–307. https://doi.org/10.1007/s00122-011-1705-6

Schwahn K, de Souza LP, Fernie AR, Tohge T (2014) Metabolomics-assisted refinement of the pathways of steroidal glycoalkaloid biosynthesis in the tomato clade. J Integr Plant Biol 56(9):864–875. https://doi.org/10.1111/jipb.12274

Scott JW, Myers JR, Boches PS, Nichols CG, Angell FF (2013) Classical genetics and traditional breeding. In: Henry RJ (ed) Genetics, genomics and breeding of tomato. CRC Press, Boca Raton, pp 37–73

Sharma A, Zhang L, Niño-Liu D, Ashrafi H, Foolad MR (2008) A *Solanum lycopersicum* x *Solanum pimpinellifolium* linkage map of tomato displaying genomic locations of R-genes, RGAs, and candidate resistance/defense-response ESTs. Int J Plant Genomics 2008:926090. https://doi.org/10.1155/2008/926090

Sharma A, Kaur M, Katnoria JK, Nagpal AK (2018) Polyphenols in food: cancer prevention and apoptosis induction. Curr Med Chem 25(36):4740–4757. https://doi.org/10.2174/0929867324666171006144208

Shirasawa K, Hirakawa H (2013) DNA marker applications to molecular genetics and genomics in tomato. Breed Sci 63(1):21–30. https://doi.org/10.1270/jsbbs.63.21

Skłodowska M, Gapínska M, Gajewska E, Gabara B (2009) Tocopherol content and enzymatic antioxidant activities in chloroplasts from NaCl-stressed tomato plants. Acta Physiol Plant 31:393–400

Slimestad R, Fossen T, Verheul MJ (2008) The flavonoids of tomatoes. J Agric Food Chem 56(7):2436–2441. https://doi.org/10.1021/jf073434n

So FV, Guthrie N, Chambers AF, Moussa M, Carroll KK (1996) Inhibition of human breast cancer cell proliferation and delay of mammary tumorigenesis by flavonoids and citrus juices. Nutr Cancer 26(2):167–181. https://doi.org/10.1080/01635589609514473

Solini A, Santini E, Ferrannini E (2006) Effect of short-term folic acid supplementation on insulin sensitivity and inflammatory markers in overweight subjects. Int J Obes 30(8):1197–1202. https://doi.org/10.1038/sj.ijo.0803265

Spencer JP, Kuhnle GG, Hajirezaei M, Mock HP, Sonnewald U, Rice-Evans C (2005) The genotypic variation of the antioxidant potential of different tomato varieties. Free Radic Res 39(9):1005–1016. https://doi.org/10.1080/10715760400022293

Stahl W, Sies H (2003) Antioxidant activity of carotenoids. Mol Asp Med 24(6):345–351. https://doi.org/10.1016/s0098-2997(03)00030-x

Stevens MA (1986) Inheritance of tomato fruit quality components. In: Janick J (ed) Plant breeding reviews, vol 4. AVI, Westport, pp 273–312

Stevens MA, Rick CM (1986) Genetics and breeding. In: Atherton JG, Rudich J (eds) The tomato crop: a scientific basis for improvement. Chapman and Hall, London, pp 35–109

Stevens R, Buret M, Duffé P, Garchery C, Baldet P, Rothan C, Causse M (2007) Candidate genes and quantitative trait loci affecting fruit ascorbic acid content in three tomato populations. Plant Physiol 143(4):1943–1953. https://doi.org/10.1104/pp.106.091413

Stommel J (2005) Sensory and objective quality attributes of beta-carotene and lycopene rich tomato fruit. J Am Soc Hortic Sci 130(130):244–251. https://doi.org/10.21273/JASHS.130.2.244

Stratil P, Klejdus B, Kubán V (2006) Determination of total content of phenolic compounds and their antioxidant activity in vegetables – evaluation of spectrophotometric methods. J Agric Food Chem 54(3):607–616. https://doi.org/10.1021/jf052334j

Sumalan RM, Ciulca SI, Poiana MA, Moigradean D, Radulov I, Negrea M, Crisan ME, Copolovici L, Sumalan RL (2020) The antioxidant profile evaluation of some tomato landraces with soil salinity tolerance correlated with high nutraceutical and functional value. Agronomy 10(4):500. https://doi.org/10.3390/agronomy10040500

Sun L, Yuan B, Zhang M, Wang L, Cui M, Wang Q, Leng P (2012) Fruit-specific RNAi-mediated suppression of SlNCED1 increases both lycopene and β-carotene contents in tomato fruit. J Exp Bot 63(8):3097–3108. https://doi.org/10.1093/jxb/ers026

Suresh BV, Roy R, Sahu K, Misra G, Chattopadhyay D (2014) Tomato genomic resources database: an integrated repository of useful tomato genomic information for basic and applied research. PLoS One 9(1):e86387. https://doi.org/10.1371/journal.pone.0086387

Takayama M, Ezura H (2015) How and why does tomato accumulate a large amount of GABA in the fruit? Front Plant Sci 6:612. https://doi.org/10.3389/fpls.2015.00612

Takayama M, Koike S, Kusano M, Matsukura C, Saito K, Ariizumi T, Ezura H (2015) Tomato glutamate decarboxylase genes SlGAD2 and SlGAD3 play key roles in regulating γ-aminobutyric acid levels in tomato (*Solanum lycopersicum*). Plant Cell Physiol 56(8): 1533–1545. https://doi.org/10.1093/pcp/pcv075

Takayama M, Matsukura C, Ariizumi T, Ezura H (2017) Activating glutamate decarboxylase activity by removing the autoinhibitory domain leads to hyper γ-aminobutyric acid (GABA) accumulation in tomato fruit. Plant Cell Rep 36:103–116. https://doi.org/10.1007/s00299-016-2061-4

Tan HL, Thomas-Ahner JM, Grainger EM, Wan L, Francis DM, Schwartz SJ, Erdman JW Jr, Clinton SK (2010) Tomato-based food products for prostate cancer prevention: what have we learned? Cancer Metastasis Rev 29(3):553–568. https://doi.org/10.1007/s10555-010-9246-z

Tanksley SD (1993) Linkage map of tomato (*Lycopersicum esculentum*) (2N=24). In: O'Brien SJ (ed) Genetic maps: locus maps of complex genomes, 6th edn. Cold Spring Harbor Press, New York, pp 39–60

Thangaraj P (2016) Proximate composition analysis. In: Pharmacological assays of plant-based natural products, Progress in drug research, vol 71. Springer, Cham

Tigchelaar EC (1986) Tomato breeding. In: Bassett JJ (ed) Breeding vegetable crops. AVI Publishing, Westport, pp 135–171

Tomlinson ML, Butelli E, Martin C, Carding SR (2017) Flavonoids from engineered tomatoes inhibit gut barrier pro-inflammatory cytokines and chemokines, via SAPK/JNK and p38 MAPK pathways. Front Nutr 4:61. https://doi.org/10.3389/fnut.2017.00061

Top O, Bar C, Ökmen B, Özer DY, Rusçuklu D, Tamer N, Doğanlar S (2014) Exploration of three *Solanum* species for improvement of antioxidant traits in tomato. HortScience 49(8): 1003–1009. https://doi.org/10.21273/HORTSCI.49.8.1003

Tousoulis D, Antoniades C, Tentolouris C, Tsioufis C, Toutouza M, Toutouzas P, Stefanadis C (2003) Effects of combined administration of vitamins C and E on reactive hyperemia and inflammatory process in chronic smokers. Atherosclerosis 170(2):261–267. https://doi.org/10.1016/S0021-9150(03)00250-8

Trujillo E, Davis C, Milner J (2006) Nutrigenomics, proteomics, metabolomics and the practice of dietetics. J Am Diet Assoc 106:403–413. https://doi.org/10.1016/j.jada.2005.12.002

Tsaniklidis G, Kotsiras A, Tsafouros A, Roussos PA, Aivalakis G, Katinakis P, Delis C (2016) Spatial and temporal distribution of genes involved in polyamine metabolism during tomato fruit development. Plant Physiol Biochem 100:27–36. https://doi.org/10.1016/j.plaphy.2016.01.001

Tsaniklidis G, Charova SN, Fanourakis D, Tsafouros A, Nikoloudakis N, Goumenaki E et al (2021) The role of temperature in mediating postharvest polyamine homeostasis in tomato fruit. Postharvest Biol Technol 179:111586. https://doi.org/10.1016/j.postharvbio.2021.111586

Uddin MS, Ferdosh S, Akanda MJH, Ghafoor K, Rukshana AH, Ali ME, Kamaruzzaman BY, Fauzi MB, Hadijah S, Shaarani S, Sarker ZI (2018) Techniques for the extraction of phytosterols and their benefits in human health: a review. Sep Sci Technol 53:2206–2223. https://doi.org/10.1080/01496395.2018.1454472

Upadhyaya P, Tyagi K, Sarma S, Tamboli V, Sreelakshmi Y, Sharma R (2017) Natural variation in folate levels among tomato (*Solanum lycopersicum*) accessions. Food Chem 217:610–619. https://doi.org/10.1016/j.foodchem.2016.09.031

USDA FoodData Central (2019) Tomato. https://fdc.nal.usda.gov/fdc-app.html#/food-details/170457/nutrients

van den Heuvel T, Renes RJ, Gremmen B, van Woerkum C, van Trijp H (2008) Consumers' images regarding genomics as a tomato breeding technology: "maybe it can provide a more tasty tomato". Euphytica 159:207–216. https://doi.org/10.1007/s10681-007-9474-7

van der Biezen EA, Brandwagt BF, van Leeuwen W, Nijkamp JJ, Hille J (1996) Identification and isolation of the FEEBLY gene from tomato by transposon tagging. Mol Gen Genet 251:267–280. https://doi.org/10.1007/BF02172517

Vats S, Bansal R, Rana N, Kumawat S, Bhatt V, Jadhav P, Kale V, Sathe A, Sonah H, Jugdaohsingh R, Sharma TR, Deshmukh R (2022) Unexplored nutritive potential of tomato to combat global malnutrition. Crit Rev Food Sci Nutr 62(4):1003–1034. https://doi.org/10.1080/10408398.2020.1832954

Verhoeyen ME, Bovy A, Collins G, Muir S, Robinson S, De Vos CHR, Colliver S (2002) Increasing antioxidant levels in tomatoes through modification of the flavonoid biosynthetic pathway. J Exp Bot 53(377):2099–2106. https://doi.org/10.1093/jxb/erf044

Wu G (2016) Dietary protein intake and human health. Food Funct 7(3):1251–1265. https://doi.org/10.1039/c5fo01530h

Wurbs D, Ruf S, Bock R (2007) Contained metabolic engineering in tomatoes by expression of carotenoid biosynthesis genes from the plastid genome. Plant J 49(2):276–288. https://doi.org/10.1111/j.1365-313X.2006.02960.x

Xia X, Cheng X, Li R, Yao J, Li Z, Cheng Y (2021) Advances in application of genome editing in tomato and recent development of genome editing technology. Theor Appl Genet 134:2727–2747. https://doi.org/10.1007/s00122-021-03874-3

Yahia EM, Contreras-Padilla M, Gonzalez-Aguilar G (2001) Ascorbic acid content in relation to ascorbic acid oxidase activity and polyamine content in tomato and bell pepper fruits during development, maturation and senescence. LWT Food Sci Technol 34(7):452–457. https://doi.org/10.1006/fstl.2001.0790

Yang S, Fresnedo-Ramírez J, Wang M, Cote L, Schweitzer P, Barba P, Takacs EM, Clark M, Luby J, Manns DC, Sacks G, Mansfield AK, Londo J, Fennell A, Gadoury D, Reisch B, Cadle-Davidson L, Sun Q (2016) A next-generation marker genotyping platform (AmpSeq) in heterozygous crops: a case study for marker-assisted selection in grapevine. Hortic Res 3:16002. https://doi.org/10.1038/hortres.2016.2

Yates HE, Frary A, Doganlar S, Frampton A, Eannetta NT, Uhlig J, Tanksley SD (2004) Comparative fine mapping of fruit quality QTLs on chromosome 4 introgressions derived from two wild tomato species. Euphytica 135:283–296. https://doi.org/10.1023/B:EUPH.0000013314.04488.87

Yazdani M, Sun Z, Yuan H, Zeng S, Thannhauser TW, Vrebalov J, Li L (2019) Ectopic expression of ORANGE promotes carotenoid accumulation and fruit development in tomato. Plant Biotechnol J 17(1):33–49. https://doi.org/10.1111/pbi.12945

Young PA, MacArthur JW (1947) Horticultural characters of tomatoes. Texas Agric Exp Station Bull 698:61

Zhang YY, Li HX, Shu WB, Zhang CJ, Ye ZB (2011a) RNA interference of a mitochondrial APX gene improves vitamin C accumulation in tomato fruit. Sci Hortic 129(2):220–226. https://doi.org/10.1016/j.scienta.2011.03.025

Zhang C, Liu J, Zhang Y, Cai X, Gong P, Zhang J, Ye Z (2011b) Overexpression of SIGMEs leads to ascorbate accumulation with enhanced oxidative stress, cold, and salt tolerance in tomato. Plant Cell Rep 30:389–398. https://doi.org/10.1007/s00299-010-0939-0

Zhang J, Zhao J, Liang Y, Zou Z (2016) Genome-wide association-mapping for fruit quality traits in tomato. Euphytica 207:439–451. https://doi.org/10.1007/s10681-015-1567-0

Zhao DK, Zhao Y, Chen SY, Kennelly EJ (2021) *Solanum* steroidal glycoalkaloids: structural diversity, biological activities, and biosynthesis. Nat Prod Rep 38(8):1423–1444. https://doi.org/10.1039/D1NP00001B

Zhou R, Wu Z, Cao X, Jiang FL (2015) Genetic diversity of cultivated and wild tomatoes revealed by morphological traits and SSR markers. Genet Mol Res 14(4):13868–13879. https://doi.org/10.4238/2015.October.29.7

Zhu Z, Chen G, Guo X, Yin W, Yu X, Hu J, Hu Z (2017) Overexpression of SlPRE2, an atypical bHLH transcription factor, affects plant morphology and fruit pigment accumulation in tomato. Sci Rep 7:5786. https://doi.org/10.1038/s41598-017-04092-y

Zou L, Li H, Ouyang B, Zhang J, Ye Z (2006a) Cloning and mapping of genes involved in tomato ascorbic acid biosynthesis and metabolism. Plant Sci 170(1):120–127. https://doi.org/10.1016/j.plantsci.2005.08.009

Zou L-P, Han-Xia LI, Ouyang B, Zhang JH, Zhi-Biao YE (2006b) Cloning, expression, and mapping of GDP-D-mannose pyrophosphorylase cDNA from tomato (*Lycopersicon esculentum*). Acta Genet Sin 33(8):757–764. https://doi.org/10.1016/S0379-4172(06)60108-X

Zouine M, Maza E, Djari A, Lauvernier M, Frasse P, Smouni A, Pirrello J, Bouzayen M (2017) TomExpress, a unified tomato RNA-Seq platform for visualization of expression data, clustering and correlation networks. Plant J 92(4):727–735. https://doi.org/10.1111/tpj.13711

Zsögön A, Čermák T, Naves ER, Notini MM, Edel KH, Weinl S, Peres LEP (2018) De novo domestication of wild tomato using genome editing. Nat Biotechnol 36:1211–1216. https://doi.org/10.1038/nbt.4272

Genome Designing for Nutritional Quality in Vegetable Brassicas

Pritam Kalia, Shrawan Singh, Raman Selvakumar, Manisha Mangal, and T. K. Nagarathna

Contents

P. Kalia (✉) · S. Singh · M. Mangal
Division of Vegetable Science, ICAR-Indian Agricultural Research Institute, New Delhi, India
e-mail: pritam.kalia@gmail.com

R. Selvakumar
Centre for Protected Cultivation Technology, ICAR-Indian Agricultural Research Institute (IARI), New Delhi, India

T. K. Nagarathna
University of Agricultural Sciences, GKVK Campus, Bengaluru, Karnataka, India

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_50

Abstract

Vegetables are rich in dietary nutrients, antioxidants, and phytoceuticals that are beneficial to health, thereby contributing to food security and ensuring nutritional security. *Brassica* vegetables are rich in dietary fiber and β-carotene, and most of them are also rich in folate, potassium, and calcium. These crops are also abundant in sulfur compounds, which vary crop by crop and provide distinctive flavor. Presence of specific glucosinolates in *Brassica* vegetables correlates with their anticarcinogenic ability. Vitamin C (cabbage and broccoli), vitamin A (broccoli), β-carotene (orange cauliflower), and anthocyanin are the most important bioactive substances found in *Brassica* vegetables (red cabbage, purple broccoli, purple cauliflower, and purple knol khol). Carotenoids serve a vital role in supporting human health, which has encouraged researchers to develop a deeper understanding of the molecular processes that regulate carotenoid production and accumulation. The *Or* gene is a unique carotenoid gene mutation that generates a high concentration of β-carotene in orange cauliflower's edible curd. The *Or* gene transformants are capable of causing carotenoids accumulation by inducing the formation of large carotenoids sequestering sheets as a carotenoid deposition sink. Anthocyanins are a kind of positively charged flavonoids prevalent in a broad spectrum of vividly colored fruits and vegetables, from orange to blue-violet. The three highest prevalent anthocyanidins are cyanidin, pelargonidin, and delphinidin. Glucosinolates are the most prominent secondary metabolites found in *Brassica* crops. Due to the vast array of health-promoting properties they possess, these glucosinolates play an important role both in plant defense and human health. Higher levels of beneficial glucosinolates (glucoraphanin and glucoiberin) and lower levels or complete elimination of progoitrin are interesting areas for quality breeding in *Brassica* vegetables. Using of association mapping, linkage disequilibrium (LD) mapping, and genome-wide association study (GWAS) was useful for

elucidating the molecular genetic foundation underpinning natural phenotypic variation. Through this several causal allele(s)/loci that had not been found in QTL mapping populations were identified. The technology of gene editing provides essential technological foundations for plant functional gene research and agricultural genetic modification. Apart from sequence analysis and assembling of DNA in *B. oleracea,* research on proteomics, transcriptomics, and metabolomics has revealed patterns of gene expression. This review also presents brief account of societal, political, and regulatory issues besides putting forth future nutritional quality improvement prospects of vegetable Brassicas.

Keywords

Cole crops · Carotenoids · Anthocyanins · Glucosinolates · Nutrients · Breeding · QTL analysis

1 Introduction

1.1 Importance of Vegetable Brassicas

Vegetables represent a large group of diverse culinary herb species and are a common component of both traditional and contemporary balanced diets. They are rich in dietary nutrients, antioxidants, and phytoceuticals that are beneficial to health, thereby contributing to food security and ensuring nutritional security. Vegetable crops played a significant role in reducing the prevalence of hidden hunger (i.e*.,* micronutrient deficiency), especially in marginal communities in vulnerable regions. The role of vegetables in maintaining the redox homeostasis of free radicals within the human body is emphasized by health professionals. This physiological process is facilitated by a variety of antioxidants and cofactors found in plant foods. However, inadequate consumption of fruits and vegetables is one of the top ten causes of the high prevalence of noncommunicable diseases (NCDs) worldwide. Low vegetable consumption is associated with anemia (iron deficiency), night blindness (vitamin A deficiency), weight loss, malnutrition-related child mortality, and susceptibility to diseases and disorders. Inadequate intake of vegetables increases chances of being overweight and obese. Vegetables are high in dietary fiber, which is essential for proper bowel movements. Brassicas are rich in glucosinolates (GSLs), which are B-thioglycoside-N-hydroxysulfates and the most abundant class of sulfur-rich secondary metabolites. Their hydrolysis products contribute to plant defense and human health. Anticancer, antibacterial, antifungal, and antioxidant properties are possessed by the GSLs. They exhibit cytotoxic and apoptotic activities that serve as cancer preventatives and reduce the risk of degenerative diseases. Inducing phase 2 enzymes by promoting anti-proliferative activity, GSLs induce phase 2 enzymes. Some glucosinolates contribute to the flavor and taste (along with sugar) of Cole crops, thereby influencing consumer preference. However, some glucosinolates, such as progoitrin, pose health risks due to

their interference with iodine absorption. To combat numerous soil-borne diseases including nematodes, the biofumigant function of GLSs is being investigated.

All of these *Brassica* vegetables are rich in dietary fiber and β-carotene, and most of them are also rich in folate, potassium, and calcium. These crops are also abundant in sulfur compounds, which vary by crop and provide distinctive flavor. These nonstarchy vegetables are regarded as healthful because of their low caloric content (about 19 calories per 12 cup raw amount) and high dietary fiber content (2.5 g per 12 cup piece). It has been shown that the presence of specific glucosinolates in *Brassica* vegetables correlates with their anticarcinogenic ability. Vitamin C (cabbage and broccoli), vitamin A (broccoli), β-carotene (orange cauliflower), and anthocyanin are the most important bioactive substances found in *Brassica* vegetables (red cabbage, purple broccoli, purple cauliflower, and purple knol khol).

Due to the presence of health-promoting antioxidants such as anthocyanins in red cabbage, purple cauliflower, purple broccoli, knol khol, and β-carotene in orange cauliflower, the various hues of *Brassica* crops attracted popular attention. In addition, *Brassica* vegetables are rich in vitamins, carbohydrates, protein, and minerals. Vegetable Brassicas possess natural variety and are capable of interbreeding, therefore there is potential to increase their nutritional value. For breeding high nutraceuticals and bioactive chemicals, such as β-carotene, anthocyanin, and glucosinolates, conventional, molecular, and genetic engineering methods are available (Kalia 2009).

The worldwide area and production of vegetable Brassicas are provided in two groupings, "cabbage and other Brassicas" and "cauliflower and broccoli," since cabbage and cauliflower are important Cole crops. The yearly output of Cole crops is 96.39 million tonnes over an area of 3.77 million hectares. Cabbage and other Brassicas contribute 70.86 million tonnes from an area of 2.53 million ha, while cauliflower and broccoli contribute 25.53 million tonnes from an area of 1.36 million ha. As indicated on worldwide distribution maps, cauliflower is cultivated in 131 nations whereas cabbage is produced in 141 nations (Fig. 1a and b). Nonetheless, China and India account for a disproportionate share of the area and production of both crops. These nations account for approximately 48.3% and 12.99% of the global cabbage and cauliflower acreage, respectively. China and India contribute nearly the same amount to global cauliflower production, with 37.4% and 34.6%, respectively. The collective proportion of these crops ranks 13th in global area and fourth in vegetable crop output. The cabbage and other Brassicas have greater average productivity (29.35 MT/ha) than world average of total vegetable crops (19.69 MT/ha). In case of cauliflower, the global average of production is 18.81 MT/ha. However, in India, the average production from these two well-known vegetable *Brassica* crops (i.e., cabbage and cauliflower) is more than the country's entire vegetable productivity (15.5 t/ha). Notably, the majority of nations have cauliflower and cabbage yields below the worldwide average.

1.2 Nutritional Composition of Vegetable Brassicas

Proximate analysis covers water, carbohydrates, protein, ash, lipid, dietary fiber, sugars, and energy value. Table 1 presents the approximate composition of *Brassica*

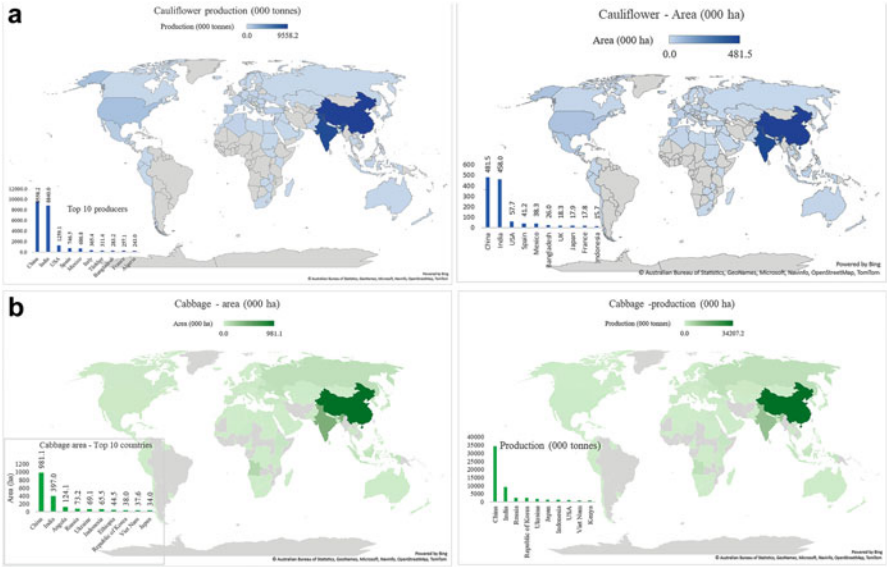


Fig. 1 (**a**) Global distribution of area and production of cauliflower and broccoli. (**b**) Global distribution of area and production of cabbage and other Brassicas. (Data source: FAOSTAT 2020)

vegetables. The vegetable Brassicas contain water in the range of 84.39–93.19%, minimum in knol khol leaves and maximum in Chinese cabbage (Table 1). Proteins are crucial macromolecules and brussels sprouts, red cabbage, collard greens, and mustard leaves are rich sources. Cauliflower curd and cabbage contain 2.15% and 1.36% proteins, respectively. Ash content refers to inorganic matter which remains after complete oxidation of organic matter and water through heating in muffle furnace. Dietary fibers are important for water retention and absorbance of organic materials in gut, thus they play role in bowl regulation, inhibition of fat accumulation, and absorbation of minerals (Jha et al. 2017). All *Brassica* vegetables have higher content of insoluble fibers than soluble fibers. Total dietary fibers in cauliflower curd portion are 3.71% and in cabbage heads it is 2.76%. Carbohydrates content and energy were low in these vegetables. Fructose is prominent sugar in Chinese cabbage, collard, cabbage, and cauliflower while sucrose in brussels sprouts and pak choi leaves.

Carotenoids content in *Brassica* vegetables are presented in Table 2. Mustard leaves, pak choi leaves, and brussels sprouts are the richest sources of lutein which contain 2939 μg, 2655 μg, and 1653 μg per 100 g of fresh weight, respectively. Similarly, β-carotene content in these crops is 2619 μg, 2450 μg, and 360 μg, respectively. Leaves of cauliflower also contain β-carotene (146 μg) but curd portion is almost devoid of it as it is just 1.59 μg/100 g of fresh weight. Among *Brassica* vegetables, mustard leaves (6397 μg/100 g fw) and pak choi leaves (5111 μg) are the richest sources while brussels sprouts (3564 μg/100 g) is top among the Cole vegetables. Red cabbage has higher content of lutein (44.50 μg), b-carotene (31.17 μg), and

Table 1 Proximate composition of *Brassica* vegetables (100 g fresh weight)

Crops	Moisture (g)	Ash (g)	Protein (g)	Total fat (g)	Dietary fiber			Carbohydrate (g) 0.04	Energy (KJ) 0.05	*Sugars*		
					Soluble (g)	Insoluble (g)	Total (g)			Fructose	Glucose	Sucrose
Brussels sprouts	84.39	1.47	4.26	0.50	0.94	3.35	4.29	0.20	0.10	0.04	0.05	0.08
Chinese cabbage	93.19	0.73	1.58	0.13	0.45	1.55	2.01	0.41	0.24	0.20	0.10	0.12
Collard greens	89.53	0.81	3.63	0.27	0.94	2.04	2.98	0.52	0.26	0.41	0.24	0.11
Cabbage	91.85	0.67	1.36	0.12	0.85	1.91	2.76	0.43	0.18	0.52	0.26	0.23
Red cabbage	91.94	0.71	1.39	0.21	0.62	1.58	2.21	0.32	0.24	0.43	0.18	0.19
Cauliflower leaves	87.64	1.22	3.90	0.42	1.06	2.37	3.43	–	0.05	0.32	0.24	0.05
Knol khol leaves	86.20	1.42	3.12	0.35	0.95	1.81	2.76	–	0.02	–	0.05	0.05
Mustard leaves	88.17	1.47	3.52	0.51	0.87	3.04	3.92	0.10	0.20	–	0.02	0.02
Pak choi leaves	93.56	1.10	1.41	0.25	0.47	1.44	1.91	0.27	0.14	0.10	0.20	0.22
Cauliflower	90.76	0.91	2.15	0.44	1.04	2.66	3.71	0.05	0.11	0.27	0.14	0.06
Knol khol	93.14	0.79	1.58	0.35	0.44	2.31	2.75	Fructose	Glucose	0.05	0.11	0.27

Source: Longvah et al. (2017)

Table 2 Carotenoids and fat-soluble vitamins in *Brassica* vegetables (μg/100 g fresh weight)

Crops	Lutein	Zeaxanthin	β-carotene	Total carotenoids	Ergocalciferol (D2)	α-tocopherol equivalent	Phylloquinones (K1)
Brussels sprouts	1653.0	35.64	360	2564	0.26	0.21	23.60
Chinese cabbage	58.00	1.50	5.50	103	0.39	0.25	111
Collard greens	143	2.68	104	358	0.18	0.20	125
Cabbage	3.98	–	20.48	273	0.21	0.05	113
Red cabbage	44.50	2.20	31.17	339	0.19	0.03	117
Cauliflower leaf	152	1.97	146	1742	4.15	0.08	144
Knol khol leaves	15.62	2.77	12.04	154.0	0.59	0.53	295
Mustard leaves	2939	8.13	2619	6397	5.40	0.57	192
Pak choi leaves	2655	5.50	2450	5111	0.10	0.03	39.85
Cauliflower	31.3	5.77	1.59	50.48	1.32	0.02	14.3
Knol khol	2.51	1.55	–	28.82	0.32	0.17	8.90

Source: Longvah et al. (2017)

total carotenoids (339 μg) than cabbage (green), i.e., 3.98, 20.48, and 273 μg/100 g fw, respectively. Ergocaciferols, α-tocopherol, and phylloquinones are important for human health, and these are reported to be high in cauliflower leaves, mustard leaves, and knol khol leaves.

Brassica vegetables are fair source of thiamin (0.01–0.8 mg per 100 g fw), riboflavin (0.05–0.22 mg), pantothenic acid (0.24–0.62 mg), pyridoxine (0.13–0.96 mg), niacin (0.21–0.86 mg), and biotin (1.08–13.57 μg) (Table 3). Not only the iron deficiency, but the prevalence of anemia is also contributed by deficiencies of folate and vitamin B12. *Brassica* vegetables are good sources of folate, particularly mustard leaves (110 μg), pak choi (98.5 μg), brussels sprouts (85.01 μg), and collard greens (63.46 μg). However, the two prominent Cole vegetables, namely cabbage (46.36 μg) and cauliflower (45.95 μg), were at intermediate level among the *Brassica* vegetables for folate content. Brussels sprouts (89.45 mg), knol khol leaves (71.11 mg), knol khol (64.7 mg), and mustard leaves (60.32 mg) were among the prominent sources of total ascorbic acid in *Brassica* vegetables.

According to the data provided by Longvah et al. (2017) and given in Table 4, *Brassica* plants are a rich source of dietary nutrients, much as other vegetables. Leafy *Brassica* vegetables contain high amount of calcium, iron, manganese, phosphorus, potassium, selenium, and zinc. The richest sources (100 g FW) among these are mustard leaves for calcium (191 mg), magnesium (66.0 mg), and selenium (10.5 μg) and pak choi for iron (3.78 mg) and zinc (0.68 mg). Potassium is highest in Brussels sprouts (639 mg) and the ratio of potassium to sodium in these crops was in the range of 7.41 in cauliflower to 34.52 in Brussels sprouts. Pak choi leaves also have a favorable K:Na ratio (21.06) followed by cabbage (15.55) in *Brassica* vegetables. These ratios are better than the recommended levels of 2.3 potassium to 1 sodium which facilitates the maintenance of normal fluid and neuronal homeostasis.

Amino acids profile of *Brassica* vegetables indicates fair content of two important sulfur-containing amino acids, namely cysteine and methionine (Table 5). Of these crops, pak choi leaves were the richest sources of cysteine (1.12 g/100 g) and methionine (1.96 g/100 g). Brussels sprout (1.34 g), cabbage green (1.06 g), and cauliflower curd (1.01 g) are also rich sources of amino acids among the prominent Cole vegetables. The *Brassica* vegetables contain all the nine essential amino acids and arginine content was present in the range of 4.15–7.63 g per 100 g. These vegetables have healthy profile of amino acids with high digestibility of protein fraction.

Oxalate is one of the antinutritional factors in *Brassica* vegetables (Table 6). In 100 g fresh edible portion, the oxalate content was highest in cauliflower leaves (172 mg) followed by Chinese cabbage (16.55 mg), pak choi leaves (14.35 mg), and brussels sprouts (12.4 mg). Insoluble oxalate has a predominant fraction in cauliflower leaves, Chinese cabbage, pak choi leaves, cabbage, and collard, while in the case of knol khol and mustard leaves the soluble fraction of oxalate is predominant. Phytate was in the range of 11.7–47.33 g in vegetable Brassicas, however, in major crops, namely cabbage and cauliflower it was estimated as 11.7 g and 18.48 g, respectively.

Table 3 Water-soluble vitamins in *Brassica* vegetables (100 g fresh weight)

Crops	Thiamine (B1) (mg)	Riboflavin (B2) (mg)	Niacin (B3) (mg)	Pantothenic acid (B5) (mg)	Pyridoxine (B6) (mg)	Biotin (B7) (μg)	Total folates (B9) (μg)	Total ascorbic acid (mg)
Brussels sprouts	0.06	0.16	0.50	0.47	0.19	2.45	85.01	89.45
Chinese cabbage	0.01	0.05	0.38	0.58	0.19	1.08	54.51	19.32
Collard greens	0.03	0.05	0.26	0.49	0.24	1.38	63.46	40.76
Cabbage	0.03	0.05	0.24	0.24	0.13	1.41	46.36	33.25
Red cabbage	0.04	0.05	0.27	0.25	0.17	1.43	34.81	43.49
Cauliflower leaves	0.05	0.05	0.21	0.34	0.23	1.38	42.99	52.84
Knol khol leaves	0.06	0.15	0.86	0.27	0.28	13.57	41.55	71.11
Mustard leaves	0.08	0.18	0.58	0.26	0.16	1.70	110	60.32
Pak choi leaves	0.02	0.22	0.66	0.31	0.96	10.25	98.50	55.60
Cauliflower	0.04	0.07	0.31	0.62	0.13	2.47	45.95	47.14
Knol khol	0.04	0.06	0.37	0.38	0.19	2.46	14.76	64.70

Source: Longvah et al. (2017)

Table 4 Mineral content in prominent *Brassica* vegetables (100 g edible portion)

Crops	Al (mg)	As (μg)	Cd (mg)	Ca (mg)	Cr (mg)	Co (mg)	Cu (mg)	Fe (mg)	Pb (mg)	Li (mg)	Mg (mg)	Mn (mg)	Hg (μg)	Mo (mg)	Ni (mg)	P (mg)	K (mg)	Se (μg)	Na (mg)	Zn (mg)
Brussels sprouts	–	0.57	0.00	53.99	0.00	20.00	10.08	1.54	0.004	–	32.99	0.41	0.06	0.013	0.012	98.56	639	2.01	18. 51	0.57
Chinese cabbage	–	–	–	58.4	60.00	–	0.05	0.39	–	–	11.51	0.19	–	–	–	33.05	258	1.85	20. 28	0.19
Collard greens	2.29		0.00	170	0.001	0.00 2	0.06	2.67	–	0.004	45.9	1.27	–	0.04	0.030	54.67	292	1.08	24.0	0.13
Cabbage	–	–	–	51.76	0.004	20.001	0.03	0.35	0.001		17.99	0.20	–	0.002	0.009	30.15	233	1.08	14.98	0.16
Red cabbage	–	–	–	48.00	0.00	40.00	20.02	0.24		0.001	26.87	0.19	–	0.002	0.004	22.14	201	1.08	24.0	0.13
Cauliflower leaves	0.46	–	0.001	96.70	0.014	–	0.14	2.42	0.006	0.003	41.5	0.50	–	0.06	0.013	62.82	374	1.05	24.31	0.31
Mustard leaves	1.74	3.65	0.004	191	0.039	0.001	0.24	2.84	0.007	0.008	66.0	0.70	0.29	0.105	1.018	55.02	309	10.50	26.0	0.42
Pak choi leaves	3.16	3.29	3.098	150.0	0.028	0.002	0.06	3.78	0.014	0.005	51.63	0.41	–	0.010	0.015	71.62	403	8.03	19.14	0.68
Cauliflower	0.17	–	–	25.16	0.007	0.001	0.05	0.96	0.004	0.001	45.28	0.36	0.08	0.02	0.024	25.95	250	0.79	33.73	0.16
Knol khol	–	–	–	35.26	0.004	0.002	0.08	0.24	–	–	23.08	0.23	–	0.002	0.013	47.33	329.5	0.47	30.72	0.31
Knol khol leaves	–	7.98	0.004	368	–	0.005	0.07	2.51	0.006	–	19.05	0.13	–	0.004	0.006	40.77	327.0	–	27.46	0.15

Source: Longvah et al. (2017)

Table 5 Amino acids profiles in *Brassica* vegetable crops (g/100 g fresh weight)

Crops	ILE	MET	HIS	CYS	PHE	TRP	LYS	VAL	ARG	THR	LEU	ALA	GLU	ASP	GLY	TYR	SER	PRO
Brussels sprouts	3.58	1.34	1.79	0.67	3.23	0.93	3.50	4.74	7.63	4.30	5.04	4.63	20.82	12.36	4.54	2.37	5.45	8.89
Chinese cabbage	5.72	1.06	1.97	0.94	3.27	1.24	5.78	5.40	5.84	4.01	5.69	6.37	21.42	10.84	3.82	2.05	5.82	3.76
Collard greens	3.81	1.11	1.62	0.91	4.56	1.24	3.62	5.64	4.32	4.04	6.67	7.68	21.91	11.13	5.61	2.78	4.20	5.35
Cabbage, green	63.56	1.06	1.56	0.90	2.56	0.93	3.12	4.83	4.57	3.76	5.31	6.26	26.30	10.55	3.24	2.11	4.02	3.98
Red cabbage	3.56	0.85	1.62	0.63	2.21	0.82	2.85	5.47	4.67	2.67	5.31	6.17	28.60	11.53	3.04	2.29	4.22	4.34
Cauliflower leaves	370	0.99	1.88	0.68	4.16	1.12	4.01	4.93	5.89	4.01	5.36	6.82	22.09	11.76	3.96	0.77	5.14	4.63
Knol khol leaves	4.31	1.37	1.96	0.95	4.93	1.05	4.27	5.70	5.30	4.45	7.20	6.76	18.00	10.33	5.31	2.90	4.48	5.39
Mustard leaves	3.22	0.74	1.91	1.02	3.53	1.31	5.42	3.52	5.74	4.08	6.57	6.60	18.67	10.86	5.31	3.21	5.11	4.65
Pak choi leaves	3.78	1.96	2.48	1.12	5.89	1.23	6.26	5.67	6.58	5.03	7.98	5.86	12.8	9.58	64.87	3.48	4.61	4.76
Cauliflower	3.84	1.01	1.80	0.56	3.88	1.06	4.13	5.81	4.15	4.02	6.01	7.96	20.70	10.61	4.15	2.89	5.15	3.45
Knol khol	3.40	0.96	0.91	0.71	1.90	0.70	2.50	5.31	5.74	3.30	4.04	5.51	36.40	9.88	3.43	2.01	4.02	3.27

Source: Longvah et al. (2017)

ILE isoleucine, *MET* methionine, *HIS* histidine, *CYS* cystine, *PHE* phenyl-alanine, *TRP* tryptophan, *LYS* lysine, *VAL* valine, *ARG* arginine, *THR* threonine, *LEU* leucine, *ALA* alanine, *GLU* glutamic acid, *ASP* aspartic acid, *GLY* glycine, *TYR* tyrosine, *SER* serine, *PRO* proline

Table 6 Oxalate and phytate content in *Brassica* vegetable crops (g/100 g fresh weight)

	Oxalate			Phytate (g)
Crops	Total (mg)	Soluble (mg)	Insoluble (mg)	
Brussels sprouts	12.4	4.67	7.73	18.32
Chinese cabbage	16.55	1.85	14.77	16.85
Collard greens	9.42	0.80	6.62	12.08
Cabbage, green	2.88	0.53	2.35	11.7
Red cabbage	2.70	0.25	2.45	12.81
Cauliflower leaves	172	10.10	162.0	13.60
Knol khol leaves	2.92	1.88	1.04	47.33
Mustard leaves	1.69	1.49	0.20	34.00
Pak choi leaves	14.35	2.15	12.20	18.05
Cauliflower	9.82	1.88	7.95	18.48
Knol khol	2.92	1.88	1.04	16.16

Source: Longvah et al. (2017)

1.3 Rising Significance Due to Long-Term Diseases and Malnutrition

Inadequate consumption of fruits and vegetables ranks among the most common causes of death from noncommunicable diseases (NCDs). The secondary metabolites in these foods contribute as antioxidants in neutralizing the free radicals and balancing the redox homeostasis. Vegetable Brassicas contain antioxidants such as vitamin C, carotenoids, and anthocyanins (purple/red), which directly act as antioxidants. They contain almost all the essential dietary minerals and amino acids, hence they have role in human nutrition. Glucosinolates in Brassicas are major health beneficial compounds. Exploiting natural mutants for new compounds, namely anthocyanins, β-carotene, and chlorophyll, through conventional breeding further enhance the scope of these crops in consumer health. These crops are good sources of Ca, Fe, and Zn, which are crucial for public health. Presence of adequate levels of vitamin C in Brassicas vegetables enhances the bioavailability of dietary minerals.

1.4 Drawbacks of Conventional Breeding and Rationale of Nutrition Genomics

Conventional breeding has played instrumental role in the creation of diverse genetic resources and varieties/hybrids in *Brassica* vegetables. However, it has limitations of handling the complex traits such as nutritional quality traits, because the success of conventional breeding for quality in *Brassica* vegetables depends upon (i) availability of genetic diversity in genepool 1 and 2 for the target trait(s); (ii) chances of linkage drag(s) or transfer of undesirable traits during breeding activities; and (iii) breeding behavior and ease of performing breeding procedure in the crop in target environment. Further, the benefit realization from improved varieties for health compounds depends

on bioavailability and response of the body. Nutrition genomics, on the other hand, describes how genetic diversity impacts the body's natural dietary intake (nutrigenetics) and also how nutrients regulate gene activity (nutrigenomics) (Guasch-Ferré et al. 2018). The term "nutrigenomics" refers to "the study of how diet may affect the expression of genetic information in an individual, and how an individual's genetic makeup affects the metabolism and response to nutrients and other bioactive components in food." This is the definition of "the study of how diet may affect the expression of genetic information in an individual" (Simopoulos 2010). It aims to (i) identify genetic variants that may be crucial for understanding genetic retorts to diet; (ii) identify genes associated with diet-related diseases; (iii) identify effective nutritional strategies to avoid or cure infections; and (iv) enhance population-level nutritional recommendations (Guasch-Ferré et al. 2018). The genetics play a key role in diet-disease relationship. Corella et al. (2013) observed that people who consumed a Mediterranean diet had improved intermediate risks for cardiovascular disease and had a reduced chance of stroke once compared to those who consumed a low-fat diet. Additionally, people who possessed the transcription factor 7–like 2 (TCF7L2) genotype that increases the chance of type 2 diabetes had a decreased probability of contracting the disease. The early stages of common illnesses and the identification of people or population groups that are at risk may assist with the appropriate design of preventative and treatment methods that are successful. The study of nutrition genomics is especially important in plant foods that are hailed as "superfoods" and recognized as important sources of certain compounds that are helpful to human health. *Brassica* vegetables are the only common source of glucosinolates which are in common use. Additionally, anthocyanin and β-carotene content are other compounds in Brassicas which have significance in human health. Thus, nutritional genomics have strong relevance for finding the target population at risk for certain dietary deficiency diseases which can be taken care of through the regular consumption of these compounds, namely glucosinolates, anthocyanin, and β-carotene, through *Brassica* vegetables. The breeding of improved genotypes of *Brassica* vegetables rich in these compounds can serve better to prevent or reduce the prevalence of target disease in human population.

2 Nutritional Elements/Compounds in Vegetable Brassicas

2.1 Chemical Composition: Content and Their Variation

2.1.1 Carotenoids in Vegetable Brassicas

Carotenoids are a category of fat-soluble chemical that gives several plant species their red, orange, and yellow colors. Because of their photoprotective and antioxidant capabilities, all photosynthetic organisms need these natural plant pigments (Lohr 2009). They protect cells from oxidative damage produced by active oxygen elements by neutralizing free radicals in the body. *De novo* synthesis in differentiated plastids of roots, flowers, fruits, and seeds is related to the dispersal of carotenoids in plants (Nagarajan et al. 2017). Chloroplasts (green plastids), chromoplasts (yellow, orange,

and red plastids), amyloplasts (plastids transporting starch), elaioplasts (plastids containing lipids), leucoplasts (colorless plastids), and etioplasts (dark matured chloroplast precursors) are all impacted by the accumulation of oxygen radicals (ROS) (Cazzonelli 2011). Carotenoids provide protection against free radical chain reactions, which may be caused by aging, smoking, pollution, and other forms of stress. A precursor of vitamin A, beta-carotene, as well as lutein and zeaxanthin has been attributed to preserving the macular degeneration from age-related neurodegeneration. In addition to protecting against macular degeneration and cataracts, carotenoids enhance eyesight and eye function. One of the "big four" vitamins, β-carotene deficiency affects billions of people, mostly women and children. Customers overwhelmingly approved of the new colors.

Carotenoids: Contents and Their Variations

The term "carotenoids" refers to a set of more than 750 naturally occurring colors that are created by plants, algae, and other photosynthetic microorganisms. These pigmented molecules are what give many different plant species their yellow, orange, and red colors. The majority of the 40–50 different carotenoid pigments present in the human diet are derived from plant foods, namely fruits and vegetables. α-carotene, beta-carotene, lycopene, lutein, β-cryptoxanthin, and zeaxanthin are the types of carotenoids that are found in foods the most often. It's possible for the body to convert the provitamin A carotenoids alpha-carotene, beta-carotene, and zeaxanthin into retinol (Fig. 1). Because they are incapable of being converted into retinol, lutein, zeaxanthin, and lycopene are classified as nonprovitamin A carotenoids (Fig. 2).

β-carotene Content in *Brassica* Vegetables

Brassica leafy greens are considered as nutritional carotenoid providers (USDA Nutrient Database 2008; Longvah et al. 2017). The carotenoid content of kale is the highest of any leafy vegetable crop, including spinach and collard greens (Sommerburg et al. 1998). Broccoli also reported as important contributor in β-carotene (1300 μg) and lutein ((1839 μg) in North American diet (Rao and Rao 2007). Among *Brassica* vegetables, cauliflower became center of attraction after identification of a natural orange mutant. The orange mutation was found in a normal, white curded autumn crop of *cv.* extra early snowball (self-compatible) growing in Bradform, Ontario, Canada, in 1972 (Crisp et al. 1975). However, the orange color in cauliflower could not attract larger section of consumers except those who could realize its nutritional importance. As a result, β-carotene-biofortified cauliflower varieties are yet to become part of diet for general public depriving them of nutritional benefit. The detailed characteristic features of the exotic parental line "EC625883," a source of *Or* gene, wild white parental line DC 18–19, an improved Indian cauliflower recipient parent, and the biofortified variety Pusa Kesari VitA-1 developed through marker-assisted backcross breeding at ICAR-IARI, New Delhi, India, are given in Table 7.

Chemical Type, Structure, and Biosynthesis Pathway of Carotenoids

Carotenoids possess main constituents that include a polyisoprenoid form, a lengthy double-bond linked network, and almost bilaterally symmetrical around the core

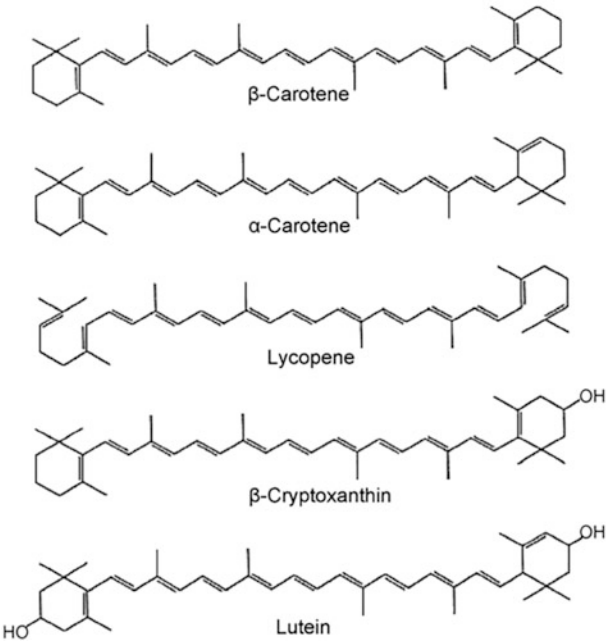


Fig. 2 Structure of some major carotenoids (Rao and Rao 2007)

Table 7 Characteristic features including β-carotene content (ppm) in white and orange cauliflower

	Cauliflower curd			
Source	DC-18-19	Pusa Kesari VitA-1 (orange*)	EC625883 (*Or* gene source genotype)	Targets for breeding program
Maturity group	Mid-late group	Mid-late group	Snowball group	
Harvesting time	Mid-December to mid-January	Mid-December to mid-January	Mid-January to March	Different maturity groups
Curd color	White	Predominantly orange	Intense orange	Orange
β-carotene content (ppm)	0.6 ppm	8–12 ppm	18–20 ppm	8–12 ppm

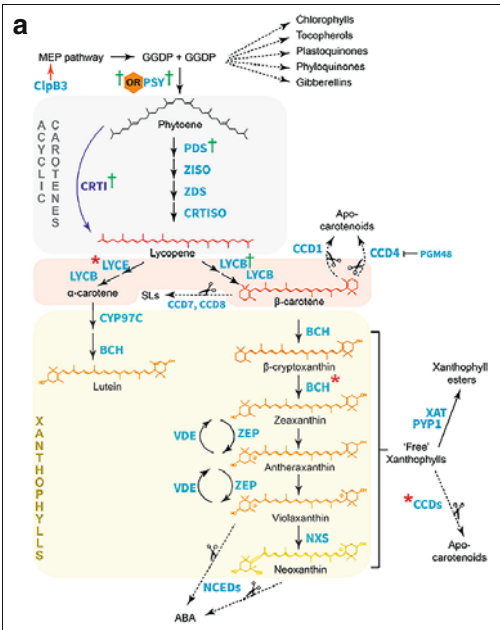
double bond (Britton 1995). Their distinct colors and antioxidant properties are the consequence of I cyclization of the end groups and (ii) the incorporation of oxygen functions into the base structure. Figure 2 depicts many typical carotenoids structures. The presence of conjugated double bonds promotes carotenoids cis-trans isomerization. In spite of the fact that trans-isomers are more prevalent and tenacious

in foods, little is established about the biological importance of carotenoid isomerization for the wellbeing of humans (Rao and Rao 2007).

In plant system, the carotenoid biosynthesis is an essential pathway for want of photosynthetic and protective pigments production, plant stress hormones, and volatile or olfactory pigments for reproductive support (Shumskaya and Wurtzel 2013; Watkins and Pogson 2020). Hence, it is considered as one of the conserved pathways in plant system, however, the accumulation of carotenoids in plants can be increased through two ways, namely: (i) altering conventional pathway as in case of most of crop plants and (ii) creating a sink area (i.e., in cell) to allow accumulation of carotenoids as in case of orange cauliflower. Thus, the present understanding about carotenoids accumulation can be explained through following four mechanisms: (i) altering biosynthesis pathway to direct synthesis for increased flux of a specific carotenoid (β-carotene or lycopene), or (ii) by raising the genes that are expressed early in the carotenoid biosynthetic pathways, substrates for specific carotenoids can be improved. (iii) Creating sink site through conversion of plastids to chromoplast which is a storage site for sequestration of β-carotene (as in case of cauliflower) and (iv) increasing the stability of produced carotenoids through changing localization within the cell or by effectively decreasing the expression enzyme(s) just below the target component known to trigger carotenoids breakdown. Researchers explained two common theories for carotenoid accumulation (or biofortification) as "push" (Sun et al. 2018) and "block" theories (Strobbe et al. 2018). According to Watkins and Pogson (2020), the "push" approach involves upregulating alleles early in the carotenoid biogenesis to drive more carbon into the beta-carotene pathways, whereas the "block" strategy involves suppressing alleles downstream of bioactive components (typically β-carotene or lycopene) or enzymes competing for the same substrate.

Carotenoids serve a vital role in supporting human health, which has encouraged researchers to develop a deeper understanding of the molecular processes that regulate carotenoid production and accumulation. Numerous details have been compiled on the genes involved in carotenoid production (Cunningham and Gantt 1998; Hirschberg 2001; Bramley 2002). Understanding the regulatory components involved in the management of carotenoid accumulation in plants is also crucial for enhancing the carotenoid status of any plant (Cunningham and Gantt 2002). This was shown by the results of Li et al. (2006), who examined carotenoid production by measuring phytoene accumulation in the presence of norflurazon, a potent phytoene desaturase inhibitor. They created calli from immature seedlings of wild-type and *Or* mutant plants (which display a high degree of carotene accumulation in tissues ordinarily devoid of carotenoids). They discovered that the calli generated from seedlings of the wild type were pale green in color. While the calli produced from *Or* seedlings had a vibrant orange hue. In addition, *Or* calli accumulated substantially more carotenoids than their wild-type counterparts. Both the wild type and *Or* calli produced considerable levels of phytoene when treated with norflurazon. The accumulation of phytoene was equivalent between the wild type and *Or* calli, and there were no significant alterations in the expression of carotenogenic genes. Their findings indicated that *Or*-induced β-carotene accumulation is not due to an increase in carotenoid biosynthetic ability.

Watkins and Pogson (2020) elaborated the sequestration mechanism as important mechanism to enhance carotenoid accumulation in plant tissues (Fig. 3a). No difference in gene amplicons from orange and white was observed by Li et al. (2006) and shown as monomorphic pattern (Fig. 3b). Nevertheless, the carotenoid-sequestering ability of the sink tissue has an effect on the number of enzymes linked with carotenoid metabolism, the chloroplast type, and the size of the storage region. The *Or* gene has an effect on the chloroplasts' response toward the source. Initially discovered in



Plant Canonical Carotenoid Biosynthesis and Regulation. GGDP-geranylgeranyl diphosphate, MEP - methylerythritol 4-phosphate (MEP) pathway, CCD-carotenoid cleavage dioxygenase, SL- strigolactone, ABA-abscisic acid, NCEDs -nine-cis-epoxy carotenoid dioxygenases, PGM48 - plastoglobule-localised metallopeptidase 48, OR- ORANGE protein. ZDS,ζ-carotene desaturase; BCH, β-carotene hydroxylase; XAT, xanthophyll acyl-transferase; CRTISO, carotenoid isomerase; CYP97C, carotene ε-hydroxylase; VDE, violaxanthin deepoxidase; ZISO, ζ-carotene isomerase; LYCB, β-lycopene cyclase; LYCE, ε-lycopene cyclase; PYP1, pale yellow petal 1. ZEP, zeaxanthin epoxidase; NXS, neoxanthin synthase; PDS, phytoene desaturase; The genes marked with a red asterisk are often targeted for downregulation in biofortification. Green represents frequent upregulated sites for carotenoid biofortification (Adapted from Watkins and Pogson, 2020).

(Adapted from Li et al. 2006).

Fig. 3 (**a** and **b**) Carotenoid biosynthesis and sequestration in cauliflower (**a**) and amplification of genes in white and orange genotypes (**b**)

cauliflower, this gene provides instructions for making a plastid-localized *DnaJ* cysteine-rich protein (2006). The *Or* gene regulates differentiation process of plastids to chromoplast in cauliflower which is essential to carotene biosynthesis. The process was also demonstrated in different β-carotene-containing melons (Lopez et al. 2008; Tzuri et al. 2015) and transgenic potatoes generated using cauliflower *Or* gene (Chayut et al. 2017; Zhou et al. 2015). Watkins and Pogson (2020) summarized the role of *Or* gene in plastid differentiation or chromoplast biogenesis is not known yet, but the mechanistic studies showed the role of *Or* in two ways: in the plastid, the *Or* protein possesses chaperone action wherein it operates by actively engaging with PSY to posttranscriptionally control its activity (Chayut et al. 2017; Welsch et al. 2018).

PSY is an important regulator and rate-limiting step in the carotenogenesis pathway. The stroma contains both an active membrane-bound and an inactive soluble form (Lätari et al. 2015; Watkins and Pogson 2020). Zhou et al. (2015) may have discovered the "*Or – PSY*" system, in which the *Or* protein interacts with PSY to improve membrane attachment and enzyme activation. It also keeps PSY folded properly, decreasing turnover through the Clp protease complex (Welsch et al. 2018). Because PSY is a rate-limiting step in numerous biological processes, this sheds light on the increase in carotene caused by *Or* genetic alteration. Furthermore, *Or* prevents additional carotenoid metabolism by inhibiting hydroxylation and/or degradation (Chayut et al. 2017).

Orange outnumbers white, although orange individuals come in a variety of shades. As a consequence of this, it would seem that the *Or* gene regulates a number of different cellular processes, having an effect on a great deal of other genes and proteins. As a result of the additional phenotypic alterations that the *Or* mutant carries, it has a pleiotropic influence on the development and growth of the plant. Plants with the dominant homozygous *Or* gene (*OrOr*) mature later and weigh much less than plants with the heterozygous *Or* gene (*Oror*), which matured later than the homozygous recessive gene (*oror*). Homozygous plants produce small, stunted curds with secondary leaves. These plants are often referred to as abnormal. The *Or* gene also promotes petiole elongation (through the *BoeRF1* genes), delays flowering, and decreases flowering shoot growth. It also gives the shoot meristem, stem pith, and vasculature at the base of the petioles an orange color.

In this pathway, plastidial methylerythritol 4-phosphate (MEP) pathway produces geranylgeranyl diphosphate (GGDP) which serves as substrate for phytoene biosynthesis (Ruiz-Sola et al. 2016). The phytoene is converted into β-carotene *via* series of desaturation and isomerization reactions (Fig. 1). In cauliflower, the pathway genes did not show any polymorphism between white and orange curding phenotypes, highlighting the conserved nature of the pathway. Here, Li et al. (2003) could establish the difference between two phenotypes, which was due to genesis of site of storage (i.e., chromoplast) in place of plastids and leucoplast in orange cauliflower. Thus, the β-carotene accumulation in orange cauliflower follows a different pathway than the conventional mechanism in other crop plants.

Plastids: Site of Carotenoids Synthesis

Carotenoids are synthesized from scratch in plastids. They are concentrated in green and nongreen tissue chloroplasts and chromoplasts, respectively. This mechanism

appears to be plant specific and substantially retained among plant species. The orange color of cauliflower is due to the existence of enormous sheet-like components in chromoplasts, but not chloroplasts (Lu et al. 2006). The orange (*Or*) gene allows curd tissue leucoplasts to convert into chromoplasts, resulting in orange curds. Plastid fusion and/or translocation factor (*BoPftf*) engages in vesicle fusion and remodels plastid internal membranes during the chloroplast-to-chromoplast transition in pepper. Figure 4 depicts the pathway of plastid-to-chromoplast conversion for β-carotene accumulation in orange cauliflower.

The *Or* gene mutation results from the insertion of a copia-like long terminal repeat retrotransposon (LTR) (4.7 kb) into exon 3 of the *Or* allele. Therefore, the *Or* mutation in cauliflower is a mutation with an obtain for β-carotene accumulation. In-frame deletion (*BoOR-Del* and *BoOR-LD*) and insertion (*BoOR-Ins*) occurred from transposition introduction into the *BoOR* region (Lu et al. 2006).

Or gene encodes a protein that is linked with the plastid and has a DnaJ cysteine-rich zinc finger motif. This protein is crucial for substantial carotenoids being deposited in tissues that are ordinarily low in coloration. *Or* has a functional role in the transformation of proplastids or other noncolored plastids into chromoplasts, providing a sink for carotenoids sequestration and storage (Zhou et al. 2008). As a result, rather than directly directing carotenoid production, the *Or* gene controls carotenoid accumulation by encouraging the growth of chromoplasts, so providing a metabolic sink to sequester and deposit carotenoids. Li et al. (2006) presented the following five evidences regarding the role of *Or* in chromoplast formation.

Lu et al. (2006) used the *Or* transgene to alter white cauliflower, revealing that Or plays a role in accelerating the conversion of proplastids and other noncolored plastids into chromoplasts for carotenoid synthesis. Furthermore, Lopez et al. (2008) discovered that the *Or* transgene expression in transgenic potatoes did not result in substantial changes at the transcript level of endogenous carotenoid biosynthetic genes and was not directly involved in regulating carotenoid synthesis. It suggested the creation of

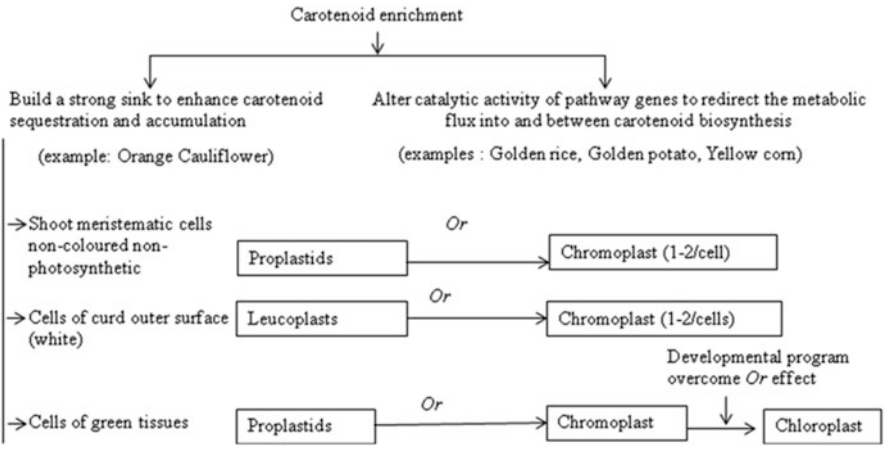


Fig. 4 Carotenoid biosynthesis processes in plants with special reference to β-carotene accumulation in orange cauliflower

chromoplasts with carotenoid sequestering structures in a heterologous system. As a result, both studies on transgenic cauliflower curd and potato tubers show that the *Or*-induced rise in carotenoid is linked to the formation of a metabolic sink.

Lopez et al. (2008) created a transgenic potato for the *Or* gene and discovered three novel metabolite intermediates of carotenoid pathways, namely phytoene, phytofluorene, and zea-carotene. Desaturation processes were restricted after the expression of the *Or* transgene. Furthermore, compared to controls, long-term cold storage boosted the carotenoid concentration of *Or* transgenic tubers tenfold. These structures were not seen in the tubers of carotenoids-rich potato types. These results show that the *Or* gene controls chromoplast differentiation and that chromoplast production regulation has a major impact on carotenoids accumulation in plants.

The *Or* protein participates in both the posttranslational regulatory oversight of phytoene synthase (PSY) that performs a frequent role in the production of carotenoids, and the maintenance of PSY proteostasis with the assistance of the plastid Clp protease network (Welsch et al. 2020). Proteostasis is the regulatory network for a balanced and functional proteome in this scenario. This network consists of competing and interrelated biological processes inside cells that govern extracellular protein production, folding, trafficking, and degradation. With 15 nuclear-encoded members and one plastid-encoded member, the Clp (caseinolytic protease) system in *Arabidopsis* is the most abundant and sophisticated soluble protease system in the plastid (Olinares et al. 2011).

Welsch et al. (2020) exploited these three differently spliced transcripts to identify three cauliflower *Or* mutant variants. Except for *BoOR-LD*, all versions produced *OR* dimers. The deletion of the first of two adjacent transmembrane domains in *BoOR-LD* inhibited *Or* dimerization and shifted the C-terminal zinc finger domain to the opposite side of the membrane. While splicing does not change the N-terminus of *BoOR*, which mediates its interaction with *PSY, BoOR* variants continue to interact with *PSY. BoOR-Del* and *BoOR-Ins* both increased BoOR protein levels in comparison to BoOR-wt, although *BoOR-Ins* did so more significantly. Furthermore, the research thus far provides the following evidence to indicate the role of the *Or* gene in carotenoids accumulation through chromoplast formation.

1. The *Or* has a considerable impact on carotenoids accumulation in proplastid- and leucoplast-rich tissues such as apical shoot meristems and the curd's outer perimeter.
2. The *ORWT* protein has been associated to nongreen plastids, and the gene is highly expressed in these tissues.
3. The existence of *Or* causes the formation of one or two large chromoplasts in each afflicted cell. It was determined that chromosomes are nothing but plastids in orange cells.
4. The *Or* mutant exhibits enhanced production of the *Pftf* gene's cauliflower homolog (*BoPftf*), which is recognized to be important in red pepper chromoplast formation.
5. Similar to the *Arabidopsis* plastid division mutant *arc6*, which has just one or two large chloroplasts in the shoot's apical meristem, *Or* halts plastid division. In

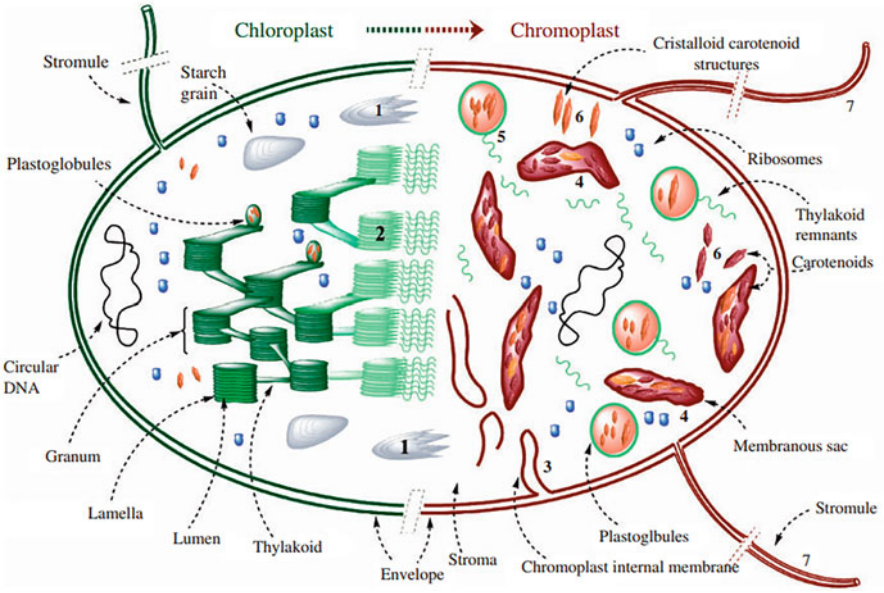


Fig. 5 The chloroplast-chromoplast transition is shown schematically. The scheme depicts the breakdown of starch granules (1) and grana and thylakoids (2); the synthesis of new membrane structures from the plastid's inner membrane envelope (3), leading to the formation of carotenoid-rich membranous sacs (4); the increase in the number and size of plastoglobules (5); the appearance of carotenoid-containing crystalloids (6); and the increase in the number of protrusions (7)

contrast to the arc6 mutant, which also has 1–2 large chloroplasts in leaf mesophyll cells, the *Or* mutant has a normal number of chloroplasts in its leaves, demonstrating that leaf growth mitigates the effects of *Or*. Figure 5 depicts the chloroplast-chromoplast transformation.

Medicinal/Physiological Properties and Functions of Carotenoids in Relation to Human Health

Carotenoids play a crucial part in the prevention of diseases as well as the upkeep of physical wellbeing. Many carotenoids, in addition to being potent antioxidants, also provide vitamin A to the body. Carotenoids' health benefits were investigated by Rao and Rao (2007). They evaluated the prophylactic effect of carotenoids in chronic disorders, as shown in Fig. 6, as well as their biological activities. Evidence supports a relationship between higher tissue concentrations of carotenoids in the human body and a reduced risk of chronic diseases, as well as a lower incidence of eye issues (Ribaya-Mercado and Blumberg 2004). It has been argued that carotenoids' antioxidant capacities are the principal method through which they exert their beneficial effects. According to Rao and Rao (2007), these phytochemicals also modulate gene expression, immune response, and Stage I and Stage II drug-metabolizing activities. β-carotene and β-cryptoxanthin have the added advantage of converting to vitamin A, which is essential in preventing disease and beginning.

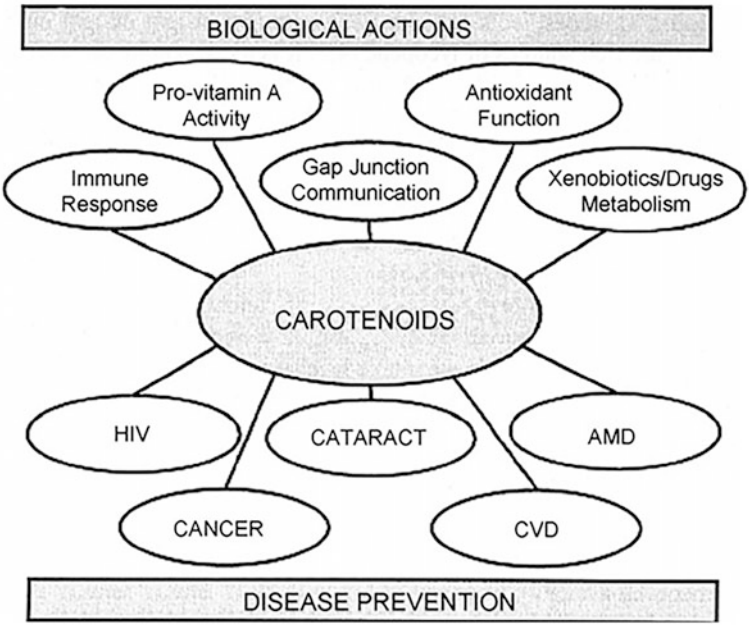


Fig. 6 Carotenoids'contribution to the fight against chronic diseases. (Adapted from Rao and Rao 2007)

However, beta-carotene increased the UVA-induced activation of hem oxygenase-1 at a dosage of 0.2 M, which indicates that it has a prooxidant impact (Obermuller-Jevic et al. 1999). In colon cancer and leukemia cells that were cultivated *in vitro*, the formation of oxygen radical (ROS) species as well as the levels of oxidized glutathione was increased when beta-carotene was present at a concentration of 10 M. Paolini et al. (2001) found that rats had a prooxidant effect and enhanced oxidative stress activity, as well as increased phase 1 enzyme activity in the liver, kidney, and gut. The Alpha-Tocopherol Beta-Carotene Cancer Prevention Study Sample (ATBCPSS) from 1994 found that supplementation with β-carotene at pharmaceutical levels led to an increased risk of lung cancer in smokers. This study also found an increased risk of death from cardiovascular disease in a sample of people who smoke, former smokers, and asbestos-exposed individuals (Omenn et al. 1996). Thus, the findings suggest that a balanced intake of β-carotene is necessary to gain the health advantages of this vitamin while avoiding the negative consequences of high amounts.

2.1.2 Anthocyanins in Vegetable Brassicas

Anthocyanins are a kind of positively charged flavonoids prevalent in a broad spectrum of vividly colored fruits and vegetables, from orange to blue-violet. The three highest prevalent anthocyanidins are cyanidin, pelargonidin, and delphinidin. Others include malvidin, peonidin, pelargonidin, and petunidin. Vegetable Brassicas

has richness of genetic resource potential for anthocyanins (Jahangir et al. 2009). Particularly red cabbage and purple cauliflower are rich sources of anthocyanins. Cauliflower and red cabbage anthocyanin profiles differed: cyanidin-3, 5-diglucoside was absent in cauliflower, but plentiful in red cabbage, coupled with the characteristic anthocyanin of the *Brassica* genus (cyanidin-3-sophoroside-5-glucoside). The p-coumaryl and feruloyl esterified forms of cyanidin-3-sophoroside-5-glucoside predominated in cauliflower, but the sinapyl ester predominated in red cabbage (Scalzo et al. 2008). Sicilian violet cauliflower landraces had less total anthocyanins, lacked cyanidin-3,5-diglucoside, and contained an isomer of cyanidin-3,5-sophoroside-5-glucoside present in red cabbage.

Anthocyanin Content

The prominent anthocyanins and reported range in *Brassica* vegetables is as below:

a) Cabbage: 1.82 mg/g FW: 36: anthocyanins: $C_{33}H_{41}O_{21}$+ (Cyanidin-3-diglucoside-5-glucose)
b) Cauliflower: 3.75 mg/g FW: $C_{30}H_{33}O_{19+}$ (cyanidin-3-(coumaryl-caffeyl) glucoside-5-(malonyl) glucoside)
c) Kohlrabi: 0.63 mg/g FW: 12: $C_{33}H_{41}O_{21}$+ (cyanidin-3-(sinapoyl)-diglucoside-5-glucoside); $C_{42}H_{47}O_{22+}$ (cyanidin-3-(caffeoyl)-p-coumaroyl-(sinapoyl) diglucoside-5-glucoside)
d) Sicilian purple: 0.48 mg/g FW: $C_{30}H_{33}O_{19+}$ (cyanidin-3-(coumaryl-caffeyl) glucoside-5-(malonyl)glucoside)

The varieties developed in vegetable Brassicas for anthocyanin content are given in Table 8.

The differences in Sicilian purple and purple cauliflower are presented hereafter (Fig. 7):

Chemical Type, Structure, and Biosynthesis Pathway of Anthocyanin

The plant anthocyanin production system is a network that has been retained in various plant species (Fig. 8). *MYB-bHLH-WD40* (MBW) is a ternary transcriptional activation complex comprising of *MYB* proteins, basic helix-loop-helix (*bHLH*) proteins, and a WD40 protein that controls structural gene expression in anthocyanin biosynthesis. Anthocyanin is governed by a single semidominant gene, *Pr*, in purple cauliflower. *Pr* expresses a tissue-specific *R2R3 MYB* transcription factor, which matches the mutant's abnormal anthocyanin accumulation pattern. Upregulation of Pr preferentially activated a *bHLH TF* and a subset of anthocyanin structural genes expressing *F3'H*, *LLDOX,* and *DFR,* to produce ectopic pigment accumulation in purple cauliflower. *Pr* gene activation is facilitated by Harbinger DNA transposon insertion into the 370 kb proximal promoter sequences of the *BoMYB2* genes. This changes the plant's phenotype. The purple sepal feature of purple broccoli is inherited by a single major gene and two secondary loci. Potential genes for the purple hypocotyl color have been discovered as *BoMYC1.1, BoMYC1.2, BoMYB114, BoTT8, BoPAL, BoDFR,* and *BoTTG1*. *BOTT8* expression was much

Table 8 Varieties developed in vegetable Brassicas for anthocyanin content

Crop	Anthocyanin content (mg/100 g FW)	Varieties of anthocyanin content	India
Cauliflower	375.0	Graffiti (3.75 mg/g FW), purple Cape, Violet Queen, Violetta Italia	KTPCF-1 (48 mg/100 g FW)
Sicilian purple	48.0	Sicilian purple	PC-1 (a line) (40 mg/100 gFW)
Cabbage	111.0.0–178.0	Premiero (1.09 mg/g), Cairo (0.8–1.53 mg/g), Red Express (1.82 mg/g), Royale (1.4 mg/g), Cardinal (0.51 mg/g), Azurro (1.44 mg/g), Bandolero (1.65 mg/g), Buscaro (1.70 mg/g)	Kinner Red Pusa Red Cabbage Hybrid-1 (7.9 mg/1gg g FW)
Purple kale	35.5	Red Dove	–
Broccoli	30	Viola, early purple sprouting, Claret	Palam Vichitra
Kohlrabi	63	Purple Vienna, Kolibri, Rapid	

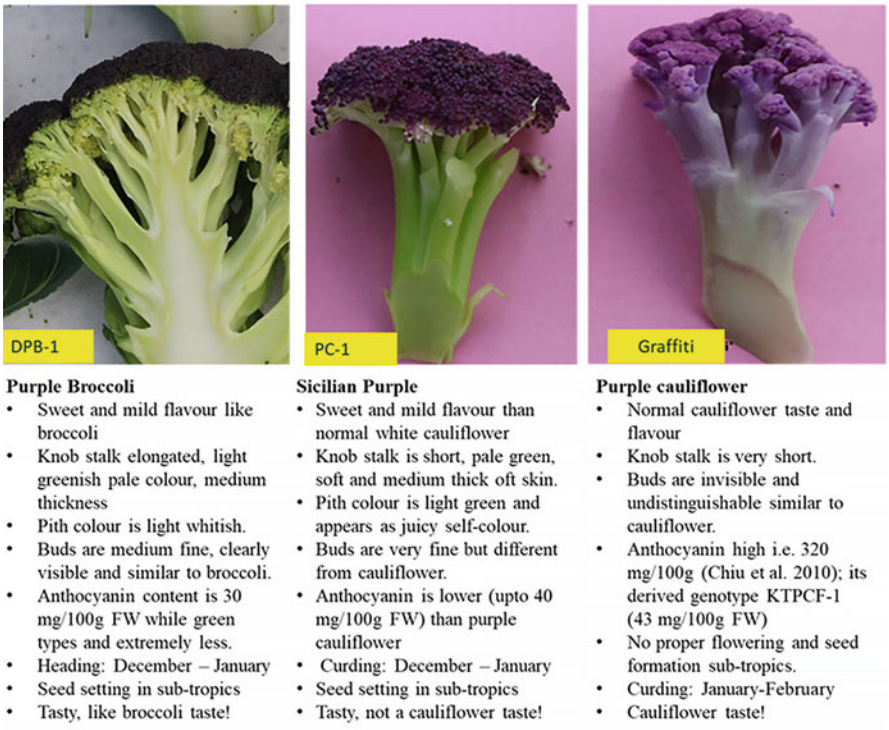


Fig. 7 Differences in purple broccoli, Sicilian purple, and purple cauliflower. DPB-1: Delhi purple broccoli, PC-1: purple cauliflower-1; Graffiti: a private sector purple cauliflower. (Modified from Singh et al. 2020)

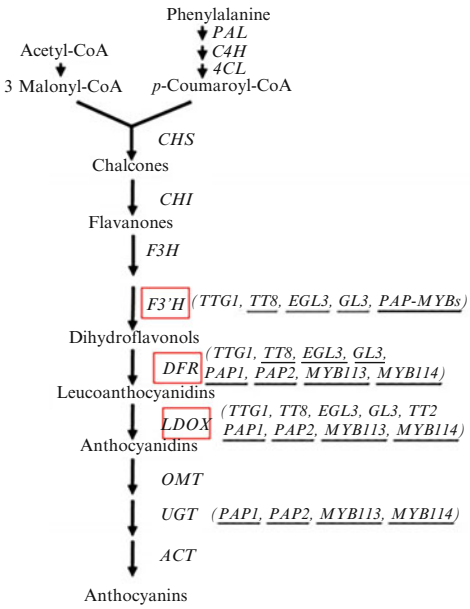


Fig. 8 Anthocyanin biosynthesis pathway in plants

greater in purple hypocotyls than in green hypocotyls. The structural genes *CHS, DFR, F3H, F3OH, GST,* and *LDOX* were constitutively upregulated at all phases of vegetative development in red types. Along with the upregulation of genes associated, the transcript levels of a bHLH gene, *BoTT8*, and a MYB transcription factor, *BoMYB2*, both went up at the same time. Furthermore, the downregulation of BoMYB3 demonstrates an equilibrium in MYB gene group components' modulation of anthocyanin production in red cabbage. The modification or deletion of the promoter of BoMYBL2-1 resulted in a purple tint in cabbage. Purple kale (*Brassica oleracea* var. *acephala* f. *tricolor*) contains total anthocyanin (1.73 mg/g), giving the mutant phenotype a vibrant purple coloration. All anthocyanin biosynthesis genes were transcribed at considerably greater quantities in the purple genotype compared to the white genotype, particularly DFR and ANS, which were hardly detectable in the white germplasm. *BoPAP1*, a *BoMYB1*, was shown to be very upregulated in purple kale and to be a necessary component for the abundant anthocyanin accumulation produced by low temperature (Zhang et al. 2012). Although *BoPAP1* was activated by light factor, the majority of anthocyanin biosynthesis genes and two transcription factors (*BoTT8* and *BoPAP2*) were considerably upregulated in the purple tissues compared to the colorless tissues in the two cultivars of knol khol. In

both cultivars' purple peel, transcript levels for late biosynthetic genes, in notably *BoF3′H, BoLDOX, BoDFR,* and *BoGST*, are significant, but hardly perceptible in other tissues (Zhang et al. 2015). Despite the expression of *BoMYBL2.2* (a negative regulator of anthocyanins) being in the opposite direction, it was discovered that *BoPAP2* and *BoTT8* were more expressed in the purple cultivar's peel than its green counterpart.

The transcriptional control of three main genes – basic helix loop helix (bHLH) proteins, WD40 proteins, and R2R3MYB transcription factors – appears to be this system's key regulatory mechanism (Broun 2005; Gonzalez et al. 2008). These genes are *bHLH* proteins (Glabra3, Enhancer of Glabra3, and Transparent Test8), *MYB113*, *MYB114*, *PAP1* (Production of Anthocyanin Pigment1), and *PAP2* (Production of Anthocyanin Pigment2) as well as the anthocyanin accumulation in *Arabidopsis* vegetative tissues (Nesi et al. 2000). The MYB transcription factor is essential for anthocyanin formation because it controls the transcription of structural genes. The *R2R-MYB* transcription factor is naturally encoded by the *Pr* gene mutation found in cauliflower, and it expresses tissue specifically (Chiu et al. 2010). They discovered that the variance in anthocyanin deposition in the 'Graffiti' cauliflower is regulated by the insertion of the Harbinger DNA transposon into the proximal promoter region of the *Pr-D* gene. (i) The kind of visible anthocyanins and (ii) the content value for case-by-case genetic investigations vary among botanical species and genotypes. MYB factors need a bHLH to control anthocyanin production in certain plants, such as *Arabidopsis* (Gonzalez et al. 2008), while in others, such as maize, they may activate transcription independently (Grotewold et al. 2000).

Medicinal/Physiological Properties and Functions of Anthocyanins in Relation to Human Health

The purple color is determined by anthocyanins, a specific type of flavonoid molecules with key biological processes in defending crops from both biotic and abiotic conditions. In addition, they contain antioxidant and anti-inflammatory characteristics that may lessen the risk of various chronic diseases in humans, including as cancer. Research indicates that anthocyanins have a role in cardiovascular disease, some forms of cancer, diabetes, disorders, osteoporosis, and allergies (Fernandes et al. 2019). Lee and Lee (2019) investigated blackcurrant extract and found it to be efficient in decreasing the pro-inflammatory M1 polarization of mouse bone marrow–derived macrophages and human THP-1 cells in response to LPS stimulation of murine macrophages. Furthermore, anthocyanins may decrease carbohydrate digestion and glucose absorption. Pancreatic lipase, β-glucosidase, and α-amylase inhibitory activity is associated with decreased gastrointestinal fat and sugar absorption. These substances combat obesity and type 2 diabetes (Castro-Acosta et al. 2016). Charron et al. (2007) examined the anthocyanin absorption rate of red cabbage in 12 volunteer groups by giving them 100, 200, and 300 g of steamed red cabbage (containing 1.38 mol anthocyanins per gram of cabbage) containing 1.38 mol anthocyanins per gram of cabbage in order to determine how well the anthocyanins were absorbed into the bloodstream. Out of 6 nonacylated and 30 acylated anthocyanins, they discovered three nonacylated, eight acylated, and

four metabolites in red cabbage. Nonacylated anthocyanins were recovered in urine at a rate four times that of acylated anthocyanins.

2.1.3 Glucosinolates in Vegetable Brassicas

Glucosinolates, also known as GSLs, are the most prominent secondary metabolites found in *Brassica* crops. Due to the vast array of health-promoting properties they possess, these glucosinolates serve an important role in both plant defense and human health. The GSLs also influence consumer preference due to their role in taste and flavor. These compounds received a lot of attention due to their perceived association with health-promoting properties (Traka and Mithen 2009; Halkier 2016). These properties are associated with *Brassica* vegetables irrespective of the kind of species, however, broccoli got special attention due to anticancer properties of "sulforaphane," a degradation product from GSLs "glucoiberin and glucoraphanin," most prominent in broccoli (Zhang et al. 1994). The breakdown products of certain specific glucosinolates possess properties which act against carcinogenesis and metastasis and also impart other health benefits through anti-inflammatory (Fahey et al. 2001; Verkerk et al. 2009), antibacterial (*Helicobacter pylori*) (Mithen et al. 2000), antifungal and antioxidant properties (Traka and Mithen 2009; Connolly et al. 2021). The GSLs reduce the risk of myocardial infraction and play role in the protection of cells from redox imbalance. These are responsible for characteristics flavor and contribute in taste (with sugar) of *Brassica* crops. Although *Brassica* vegetables also contain other health beneficial constituents, such as carotenoids, minerals, phenolics, and vitamins, but their predominant biological activity is due to glucosinolates. Some glucosinolates in oilseed Brassicas behave as growth inhibitory factors and considered as antinutritional factors (Mailer et al. 2008). Sinigrin and progoitrin are also responsible for the bitter flavor of the crucifers which influence consumers' acceptance and affect public health (particularly progoitrin). Vegetable Brassicas are known for their distinctive flavor, which is due in part to the metabolic by-products of the chemicals; however, some studies indicate little correlation between glucosinolates and flavor as in the case of broccoli (Baik et al. 2003). The GSLs are useful for soil biofumigation against various soilborne pathogens and nematodes.

These breakdown products can include isothiocyanates, thiocyanates, nitriles, epithionitriles, and oxazolidines depending on the pH levels, the presence of ferrous ions, the amount and activity of epithiospecifier protein, and the substrate. Other possible breakdown products include thiocyanates, nitriles, and epithionitriles (ESP). At neutral pH, the primary degradation component of GSL is isothiocyanates, however, at lower pH, nitriles predominate. Thiocyanates are synthesized from allyl-, benzyl-, and 4-(methylthio) butyl GSLs. There is a possibility that these substances act as inhibitors of iodine absorption in the thyroid, and elevated levels may be associated with adverse consequences on the liver (Traka 2016). The alkenyl GSLs give rise to epithioalkanes in the presence of epithiospecifier protein (Lambrix et al. 2001). Myrosinase and epithiospecifier proteins, in particular, have a lower thermal tolerance (70 and 60 degrees Celsius, respectively) (Matusheski et al. 2004). Thus, the cooking processes have been extensively researched to maximize the higher recovery of ITCs (Bongoni et al. 2014; Dosz and Jeffery 2013; Rungapamestry et al. 2006).

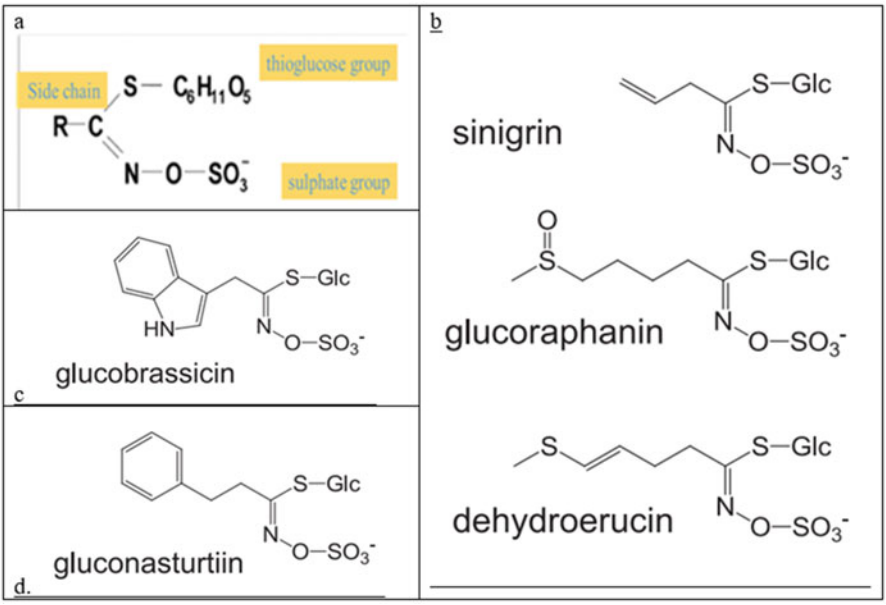


Fig. 9 Basic chemical structure of glucosinolates (**a**), indole (**b**), aromatic (**c**) and aliphatic (**d**)

The glucosinolates have a variable side chain (R), β-thioglucose moiety, and a sulfonated oxime moiety, that is produced from an amino acid as its constituent parts (Fig. 9). Basic structure of GLSs comprises a glucose residue, a sulfate group, and a variable aglycone. The sulfate group is normally balanced by a (potassium) cation. The side chain R (amino acid derived) determines whether the GLS is aliphatic (nonaromatic), indole (very stable, cyclic), and aromatic. Giamoustaris and Mithen (1996) postulated that the indolic GSLs from tryptophan, aliphatic GSLs are originated from methionine and aromatic GSLs derived from phenylalanine or tyrosine. All four of these amino acids – alanine, leucine, isoleucine, and valine – are precursors to aliphatic glucosinolates. In *Brassica* plants, methionine-derived glucosinolates are the most common (making up about half of all glucosinolates), and the breakdown products of these molecules have been shown to have metabolic activity (Mithen 2003). However, the biological activity of phenylethyl- and benzyl-GSLs (aromatic GSL breakdown metabolites) and indole GSLs have also been documented. In 1956, the compositions of sinalbin and sinigrin were the initial glucosinolates to be determined (Ettlinger and Lundeen 1956). In 1961, the word "glucosinolate" was introduced for the first time (Ettlinger and Kjaer 1968).

Glucosinolates

In 16 dicotyledonous plant families, about 136 distinct GLs structures have been identified (Clarke 2010; Agerbirk and Olsen 2012). Glucosinolates are prevalent in plants because the sugar component is easily modified by a variety of side chain modifications, *viz*. alkenylation, hydroxylation, S-oxygenation, etc. and acyl

conjugation (Halkier 2016). There are significant quantities of up to four unique glucosinolates in every plant species (Fahey et al. 2001). Investigating families other than the Brassicaceae regularly yields new data, but there is no database that tracks these studies. Bennett et al. (2004) have studied Brassicaceae plants previously.

The glucosinolate concentration of a number of *Brassica* vegetable tissues is approximately 1% of dry weight (DW), whereas it has been observed to reach 10% in the seeds. The discovery of large amounts of glucoraphanin (4-[methylsulfinyl] butyl isothiocyanate) in broccoli drew attention among cultivated Brassicas. The concentrations of the most prevalent glucosinolates are listed in Table 9. Brassicas contain the glucosinolates isothiocyanate (cauliflower, cabbage, and knol khol), sulforaphane (broccoli – most essential in cancer prevention), glucobrassicin (cabbage and broccoli), gluconasturtiin (cabbage), indole-3-carbinol (cabbage), and sinigrin (brussels sprout and broccoli). 4-(methylthio) butyl glucosinolate, 3-(methylthio) propylglucosinolate, and 2-phenethylglucosinolate were the predominant glucosinolates in kohlrabi (Macleod and Macleod 1990). Breeding cultivars, which produce a large amount of glucosinolates, lead to the production of sulforaphane compared to those that produce nitriles. Introgressed *Brassica villosa* chromosome segments led to an increase in glucoiberin levels. The introgressed region also determined whether hydrolysis produces iberin or sulforaphane. These findings suggest that broccoli with a high concentration of glucosinolates may be beneficial for increasing consumption.

Table 9 Major glucosinolates (GLSs) in commercial *Brassica* vegetables

Crucifer vegetables	Glucosinolates groups	Aliphatic glucosinolates (GLS) Range (mg/100 g fw)
Green broccoli	Glucoraphanin	11.6–34
Purple broccoli	Glucoraphanin	6.7
White cauliflower	Glucoiberin	0.5–6.6
Green cauliflower	Glucoiberin	1.2–27.7
Purple cauliflower	Glucoraphanin	11.6
	Glucoiberin	4.6
Red cabbage	Glucoraphanin	4.0–18.2
	Glucoiberin	4.0–13.6
	Sinigrin	3.0–16.7
Savoy cabbage	Glucoiberin	10.4–21.2
	Sinigrin	15.5–18.6
Brussels sprouts	Sinigrin	22.0–25.3
	Glucoiberin	6.4–13.9
Kale	Sinigrin	2.2–22.7
	Glucoiberin	13.4–16.0

Adapted from Verkerk et al. (2009)

Glucosinolate Content in Some Tissues

Fahey et al. (2001) collated information on GSLs for a range of species, but did not provide numerical data. Brassicaceae family members of high commercial value, including as cauliflower, broccoli, cabbage, brussels sprouts, and root crops like as radish and turnip, are the GSL domesticated plants that have been studied the most thoroughly. McNaughton and Marks (2003) and Verkerk et al. (2009) established a database including the total GSLs of raw, frozen, boiling, and cooked cruciferous vegetables. The principal GSLs, GSL content, and influence of cruciferous vegetables on human health were examined. Bennett et al. (2004) undertook an exhaustive examination of the GSLs found in seed. Based on GSL content, our exhaustive screening of seeds revealed the following categories:

(i) Arylaliphatic GSLs characterized by widespread substitution (such as 3,4-dimethoxybenzyl, 3,4,5-trimethoxybenzyl GLs and 3, 4-dihydroxybenzyl).
(ii) There are only aliphatic GLs with lengthy chain lengths.
(iii) Only aliphatic GLs with shorter (C-3) to medium (C-4) chain lengths are allowed.
(iv) There can only be simple arylaliphatic GSLs (2-phenylethyl (21), benzyl (20), and 4-hydroxybenzyl (22), GSLs).

The prominent glucosinolates in *Brassica* vegetables are mentioned in Fig. 10.

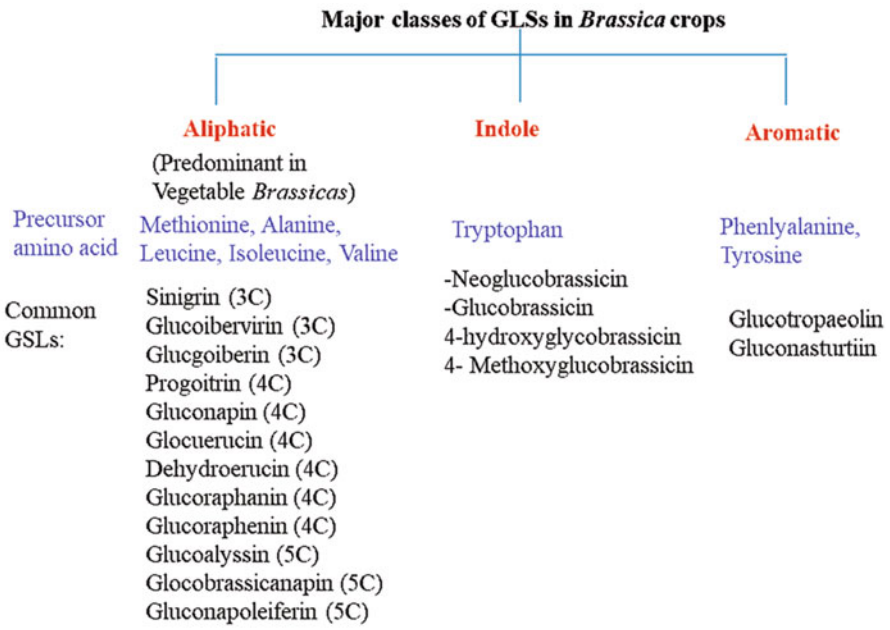


Fig. 10 Major glucosinolates in *Brassica* vegetables

Factors Determining Glucosinolates Content

The GL content of plant cultivars, individual plants, and plant parts is controlled by genetics, the environment, and plant nutrients. The greatest concentration of GLs is found in the plant's juvenile tissues. Identification and determination of phytochemical composition in plant tissues such as seed and leaf give a full understanding of these substances' potent bioactive application in agriculture as potentially insecticides and as functional nutraceuticals in animal and human diets (Berhow et al. 2013). Specialized extraction methods may be capable of separating a few GLs from acceptable plant material, but chemical synthesis is often required to get large amounts of GLs (Rollin and Tatibouët 2011). These variables impact the variation of glucosinolates:

i. Plant species: Each plant species in the Brassicaceae family has a distinct GLS profile made up of major GLSs. It also varies across genotypes within a species. Broccoli GLS contributors include *Brassica villosa*, *Brassica atlantica*, and *Brassica drepanensis*.
ii. Broccoli GLS include: Glucoraphanin: 2.2–18.4 mmol g^{-1} Aliphatic: 3.0–24.1 mmol g^{-1} Indolic: 1.0–4.9 mmol g^{-1}.
iii. Plant organ or part: Although the great majority of GLSs are present in all plant organs, the concentration and content of GLSs may vary significantly and alter during plant development. The concentration of GLSs in the roots of many Brassicas was higher than in the shoots.
iv. 4-mercaptobutylglucosinolate and 4-methylsulfinylbutylglucosinolate are found in the leaves and flowers of *Eruca* species, respectively.
v. Plant age: From transplanting to harvesting, the GLS concentration in pak choi and potherb mustard dropped. Sinigrin levels decreased from seedling to early flowering, increased during late flowering, and then decreased until seed maturity. It was higher in immature, less developed broccoli heads than in well-established broccoli heads due to a decrease in indole GLSs (Fig. 11).
vi. Plant development stages: GLS concentration was highest in the second development stage (42 DAT) with low sulfur fertilization and lowest in the third development stage with rich sulfur fertilization; afterward, GLS concentration decreased until the overmaturation stage. Sulforaphane (SFN), a GLS glucoraphanin breakdown product, flourished in broccoli until the seventh commercial maturity stage.
vii. Other (nongenetic) variables include: Soil, climate, damage, and fertilization, as well as determination processes, and so on.

Chemical Type, Structure, and Biosynthesis Pathway of Glucosinolates

The GSLs are most investigated pathways and represent a successful example of introgression of genomic regions (QTLs) for biofortification of beneficial GSLs (glucoraphanin and glucoiberin) in broccoli from a wild species *Brassica villosa* (Faulkner et al. 1998; Sarikamis et al. 2006). It is a genetically complex trait governed by several pathway genes and phenotyping of the population is also difficult and costly process. Although, some sensory attributes are related with glucosinolates content (Baik et al. 2003) but this evidence lacks strong correlation for use in breeding programs. Thus, use of alternative but robust tools such as

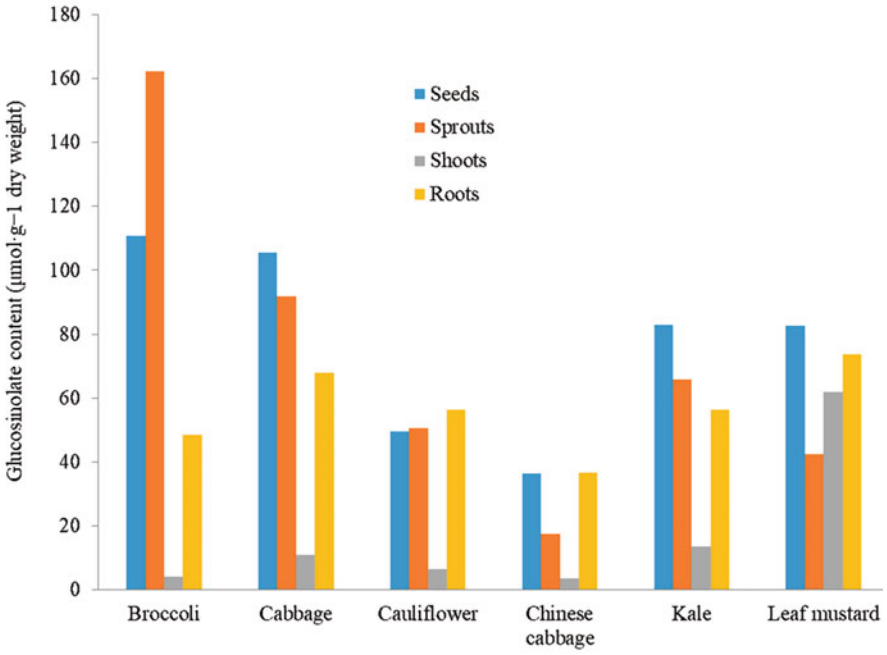


Fig. 11 Variation in GSLs content in Cole crops and their parts. (Data source: Bhandari et al. 2015)

molecular tools are more promising in tracking the introgression of gene(s) for desirable glucosinolates in *Brassica* vegetables.

Chemical Type, Structure, and Biochemical Pathways of Glucosinolates Production

The building blocks of glucosinolates are a variable side chain (R), a sulfonated oxime moiety, and a β-thioglucose moiety originating from an amino group (Fig. 12). Basic structure of GLSs comprises a glucose residue, a sulfate group, and a variable aglycone. The sulfate group is normally balanced by a (potassium) cation. The side chain R (amino acid derived) determines whether the GLS is aliphatic (nonaromatic), indole (very stable, cyclic), and aromatic. The aromatic GSLs derived from tyrosine or phenylalanine and aliphatic GSLs are derived from methionine, indolic GSLs from tryptophan (Giamoustaris and Mithen 1996). Aliphatic glucosinolates are also derived from isoleucine, leucine, valine, and alanine. Most glucosinolates in *Brassica* plants come from the amino acid methionine (nearly half of all glucosinolates), and their bioactivity is dependent on the metabolic by-products (Mithen 2003). However, the biological functions of oxidation of aromatic GSLs such as benzyl and phenylethyl GSLs, as well as indole GSLs, have also been documented. In 1956, the properties of sinigrin and sinalbin, two glucosinolates, were determined to be the first to be understood (Ettlinger and Lundeen 1956). In 1961, the word "glucosinolate" was coined for the first occurrence (Ettlinger and Kjaer 1968).

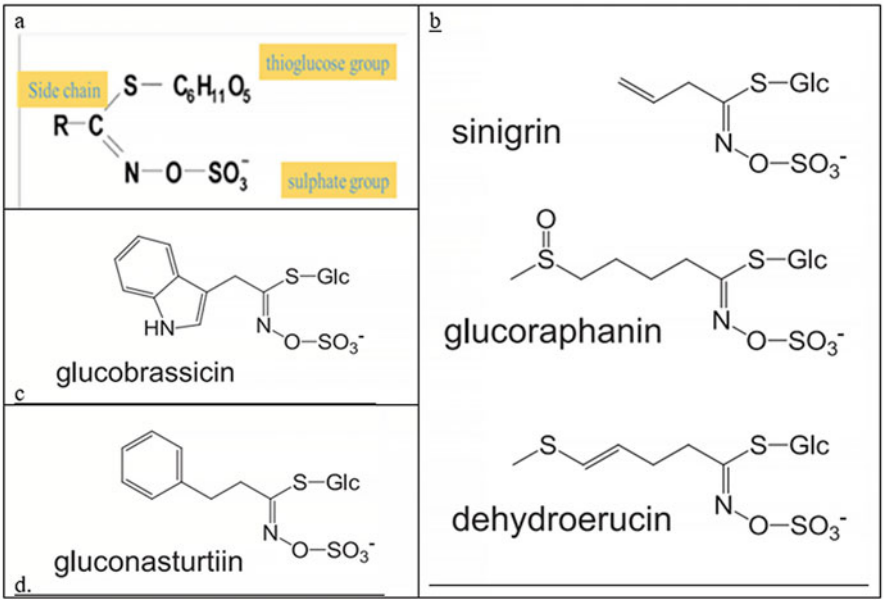


Fig. 12 Basic chemical structure of glucosinolates (**a**), indole (**b**), aromatic (**c**), and aliphatic (**d**)

Glucosinolate Biosynthesis in *Brassica* Vegetables

Extensive study has been conducted on the formation of glucosinolates, mostly employing *Arabidopsis* (Grubb and Abel 2006). Based on the precursor amino acid, glucosinolates are biosynthesized from aliphatic (mostly methionine), indole (from tryptophan), and aromatic amino acids (phenylalanine or tyrosine). Three distinct processes compose the glucosinolate synthesis routes: I chain elongation, (ii) assembly of a fundamental glucosinolate molecule, and (iii) alteration in the second stage (Fig. 13) (Sønderby et al. 2010). Ishida and collegues (2016) detailed the method for generating GSL from Brassicaceae plants. Aliphatic GSLs contain many side chains of varying lengths that are determined by periods of chain elongation. First, branched-chain amino acid aminotransferase transforms amino acids like methionine into the necessary 2-oxo acids for chain elongation (BACT). These 2-oxo acids serve as precursors in the elongation of methylene groups. The elongation is caused by the enzymes isopropylmalate isomerase (IPMI), isopropylmalate dehydrogenase (IPMDH), and methylthioalkylmalate synthase (MAM). BCAT converts the extended 2-oxo acids into their respective amino acids.

Medicinal/Physiological Properties and Functions of Glucosinolates in Relation to Human Health

The structural feature retained by intact glucosinolates and diverse breakdown products is seen in Fig. 14. Traka (2016) analyzed the results of epidemiological research on glucosinolates. Consumption of vegetable Brassicas, which are an important source of glucosinolates, has been linked to a considerably lower risk

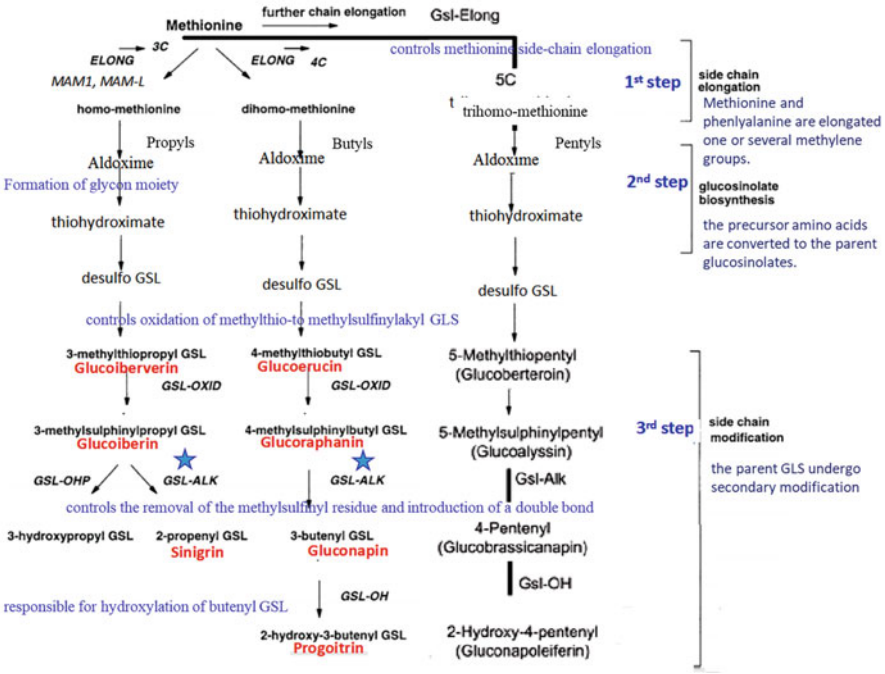


Fig. 13 Biosynthesis of glucosinolates in *Brassica* vegetables

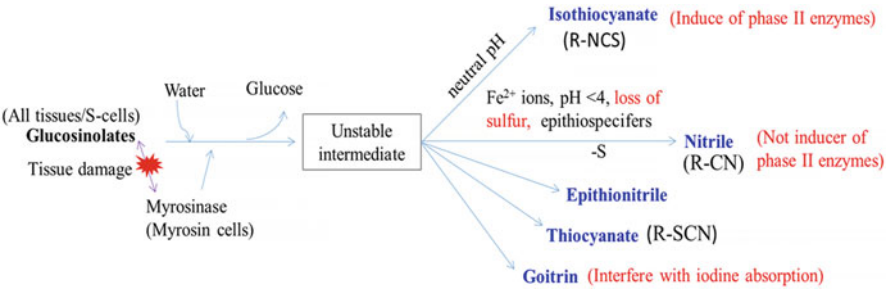


Fig. 14 Glucosinolates-myrosinase system (aliphatic GSL)

of cardiovascular disease (Giovannucci et al. 2003; Spitz et al. 2000), stomach (London et al. 2000, lung , prostate (Hansson et al. 1993; Seow et al. 2002), and rectal cancers. Richman et al. (2012) demonstrated that ingesting foods rich in glucosinolates, such as broccoli, is involved with a decreased chance of advancing from a less aggressive type of prostate cancer to a more aggressive cancer of the disease (Halkier and Gershenzon 2006). This implies that glucosinolates contribute to health improvement even after the onset of sickness. The hydrolytic product of glucosinolate and metabolic products operate as chemoprotective agents against chemically induced carcinogens by inhibiting the beginning of

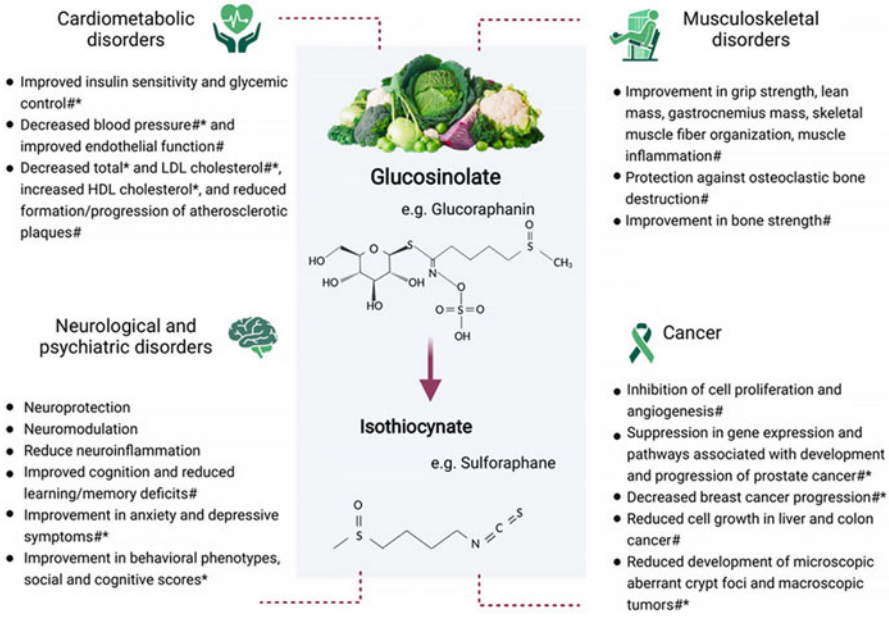


Fig. 15 Health benefits of glucosinolates from vegetable Brassicas. (Adapted from Connolly et al. 2021)

tumors in several rodent organs, including the liver, colon, mammary gland, pancreas, etc. (Fig. 15).

Suforaphanes from *Brassica* Vegetables

The cytotoxic and apoptotic function of GSLs in damaged cells acts as a cancer preventative and reduces the incidence of degenerative disorders. By stimulating antiproliferative action, they activate phase 2 enzymes. The GSLs 3-methylsulfinylpropyl (glucoiberin) and 4-methylsulfinylbutyl (glucoraphanin) are present in *Brassica* crops, with broccoli having the highest concentration (Fig. 16). The lack of a functional allele at the *GSL-ALK* gene is responsible for this. Sulforaphane is the most effective GSL breakdown product derived from glucoraphanin, which stimulates phase 2 enzyme and is believed to have a significant role in the detoxification of xenobiotic substances when ingested with food (Fahey et al. 2001; Rosa et al. 1997). Sulforaphane is the most effective naturally occurring inducer of phase 2 enzymes. It is implicated in apoptosis activation, cell cycle arrest induction, and NFB inhibition (Tortorella et al. 2015). In addition, they hypothesized that sulforaphane may target epigenetic changes in some cancers, correcting abnormal alterations in transcriptional activation through global demethylation, microRNA manipulation, and histone deacetylase suppression.

Evidence Suggesting Role of Isothiocyanates in Human Health

The breakdown product sulforaphane is important because it induces phase 2 (detoxification) enzymes *viz.* UDP-glucuronosyl transferases, glutathione-S-transferases,

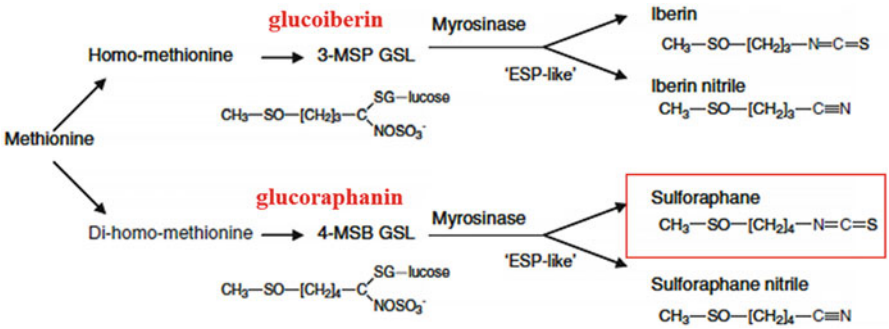


Fig. 16 Structure and degradation of 3-methylsulfinylpropyl (3-MSP) and 4-methylsulfinylbutyl (4-MSB) glucosinolate. The ratio of the two glucosinolates is determined by the extent of methionine elongation, determined by genes within the introgressed segments on O_2 and O_5. The ratio of isothiocyanates and nitriles produced depends upon the activity ESP. (Adapted from Sarikamis et al. 2006)

and quinone reductase, *via* ARE-mediated transcription which play role in neutralizing toxic electrophilic metabolites of carcinogens and ROS and serve as anticarcinogens early in the process of carcinogenesis. The cell cycle of many different kinds of cells, including prostate, lyphocyte, colon, and mammary, can be inhibited by sulforaphane's availability. It induces apoptosis in a range of cell lines. The mechanism implicated in apoptosis is JNK (c-Jun NH (2)-terminal kinase) signal transduction pathway and activation of caspases (a family of protease enzymes play essential roles in programmed cell death). Sulforaphane and synthetic cyclic isothiocyanate analogues block mammary tumor development in Sprague-Dawley rats treated with 9, 1O-dimethyl-1,2-benzanthracene (DMBA) (Huggins et al. 1961; Zhang et al. 1994; Welsch 1985). Methylsulfilalkyl isothiocyanates are known to be effective causative agents of phase 2 detoxifying enzymes in murina hepatoma *Hepa1c1c7* cells in culture. These enzymes are related with a decreased sensitivity of humans and mammalian tissue samples to the toxic and neoplastic effects of carcinogens (Faulkner et al. 1998). Administration of sulforaphane (75 or 150 μmol per day for 5 days) in Sprague-Dawley rats treated with 9,10-dimethyl-1,2-benzanthracene (carcinogen) around the time of exposure the incidence, multiplicity, and weight of mammary tumors were significantly reduced, and their development was delayed (Zhang et al. 1994).

Myrosinase: Enzyme for GSL Breakdown

The GSLs coexist with myrosinase (thioglycosidase) in a system called "glucosinolate-myrosinase," but they are kept physically apart in plants. The GSLs are believed to be mostly stored in the vacuoles of S cells and are water soluble. The S cells are referred to as sulfur-rich cells because their cell sap contains up to 350 mM S, with GLS accounting for 84% of the sulfur, or around 146 mM GSLs. It should be noted that while the GSLs are thought to be biologically inert chemicals by themselves, mechanical injury, infection, or insect assault cause the GSLs to be hydrolyzed, which produces a variety of physiologically active molecules (Bones

and Rossiter 1996). Myrosinase catalyzes the hydrolysis, resulting in the breaking of the thio-glucose link and the production of unstable thiohydrosimate-O-sulfonate, which is then rearranged to form a number of compounds (Bones and Rossiter 1996). It primarily generates isothiocynate (ITC), which controls the bioactivity of *Brassica* plants, particularly their anticancer capabilities. Other products of the hydrolysis include thiocynates, nitriles, and elemental sulfur.

Plants have the enzyme myrosinase, which, in the presence of water, separates the glucose group from a glucosinolate. When the remaining molecule rapidly converts into an isothiocyanate, nitrite, or thiocyanate, the active components are produced (Traka et al. 2008). Some of these breakdown products include isothiocyanates, thiocyanates, epithionitriles, epithioalkanes, nitriles, and oxazolidine-2-thiones (Grubb and Abel 2006). At a higher acidic pH, nitriles are produced, while isothiocyanates predominate at physiological pH (Halkier and Du 1997). It has long been known that some glucosinolate breakdown products used as feed and food components offer special benefits for plant defense and human nutrition. This characteristic has led some people to consider *Brassica* plants as possible functional foods. "Functional foods" are defined as those that, as regular dietary components, may provide enough quantities of bioactive components that are advantageous for improving health (Carlson et al. 1987b; VanEtten et al. 1976). Studies evaluating the glucosinolate proficiencies of cruciferous crops reveal considerable differences (Ciska et al. 2000). Cartea and Velasco (2008) previously posted an overview of the major glucosinolates discovered in common vegetable Brassicas, *viz*. *B. oleracea*, *B. napus*, and *B. rapa*. In *Brassica* plants, it has been discovered that the amounts of indole and aliphatic glucosinolates, as well as the glucosinolate profile, alter significantly according to the developmental phase, biological makeup, and external conditions (Velasco et al. 2007). Numerous epidemiological investigations have shown a substantial link between diet and cancer. Numerous research conducted over the last 30 years have shown that naturally occurring substances like glucosinolates, which have anticancer properties, are present in fruits and vegetables (Rosa et al. 1997; Talalay and Fahey 2001; Block et al. 1992; Smith et al. 2005). As a result, interest in cruciferous vegetables for their potential role in treating cancer has increased over the past several years (Anilakumar et al. 2006; Farnham et al. 2004). Numerous studies have shown that the chemopreventive properties of the *Brassica* vegetable family have an effect on both the initiation and advancement stages of the carcinogenic process. This concept is supported by similar findings from clinical trials and epidemiological studies (Zhang and Talalay 1994). Isothiocyanates and indoles are two important subclasses of the autolytic degradation products of glucosinolates with anticancer properties. These substances have been attributed to both the stimulation of detoxifying activities (phase II enzymes) and the inhibition of initiation enzymes in in *vivo* and *in vitro* studies of the origins of cancer (phase I enzymes) (Fahey et al. 2002). Glucosinolates are sulfur compounds that give these plants their distinctive flavor and odor in addition to their medicinal properties. Although the relationship between glucosinolate content and sensory characteristics is complex, isothiocyanates produced from glucosinolates give *Brassica* products their bitter, pungent flavor and sulfurous aroma (Padilla et al. 2007).

Myrosinases are, thus, glycoproteins that are thought to be produced in the rough endoplasmic reticulum, glycosylated in the Golgi apparatus, and then stored in the vacuoles (or myrosin grains) of myrosin cells. The thioglucoside glucohydrolases known as myrosinases catalyze the hydrolysis of the thioglucosidic link. Together with other glycosidases, they belong to the glycoside hydrolase family 1, which is involved in plant defense. Two identical, extensively glycosylated 55–65 kDa polypeptides make up the myrosinase. Two conserved residues from each monomer tetrahedrally coordinate a $Zn2^{+}$ to maintain the dimeric structure (Fig. 15).

Common properties of myrosinase are:

(i) Resistance to heat (temperature optimality up to 70 degrees Celsius).
(ii) The activation of these compounds by modest concentrations of ascorbic acid (1–2 mM).
(iii) Depending on the species, the optimal pH range might be either wide or tight.
(iv) Differ in their affinities for the various GLS as well as the efficiencies with which they convert.
(v) Species that have GLS might have a variety of myrosinase isoforms, each of which could have unique enzymatic activity as well as distinct patterns of temporal and spatial expressions.

Evidence for Role of GSLs in Human Health

In the human body, GSLs are hydrolyzed into ICTs at two sites: (i) in the mouth by plant myrosinase and (ii) in the human distal intestine by myrosinases generated by intestinal bacteria. Passive diffusion absorbs the ITCs. The glutamine in ITCs conjugates naturally and is enhanced by glutathione-S-transferases such as Mu (*GSTM1*), which is considered the key determinant of ITC metabolism, *Theta* (GSTT), and *Pi* (GSTP1) (Hayes et al. 2008). With the aid of MRP1 and MRP2 (multidrug resistance–associated proteins) and p-glycoproteins (Pgp-1), the glutathione component is rapidly released extracellularly. After that, the metabolites of the ITC are delivered to the liver via the liver's portal vein so that they may be processed by the mercapturic acid process. Following that, enzymatic changes such as glutamine and glycine cleavage generate cysteine-glycine conjugates and cysteine conjugates, which are then acetylated to form N-acetyl-cysteine (NAC) conjugates that are eliminated in urine, as detailed by Brüsewitz et al. (1977). Traka (2016) revealed various mechanisms by which free ITCs and their compounds are absorbed by hepatocytes and transported to the plasma systemic circulation. In the blood systemic circulation, they may enter periphery organs at lower quantities than in the liver (Traka 2016). The metabolic fate of GSLs in humans after consuming broccoli has been the subject of a significant amount of research. Epidemiological studies have shown that the preventative benefits of eating crucifers may vary owing to variations in glutathione S-transferases (GSTs). There are several polymorphisms in the genes that belong to the GSTs family, including null mutations in GSTM1 and GSTT1, which result in the lack of a gene product that is functional (Traka 2016). Cotton et al. (2000) discovered differences in *GSTM1* homozygous null genotype rates ranging from 30–63%, whereas *GSTT1* homozygous null genotype rates varied from 10–21% for Caucasians and up to 64% for certain Asian races. The

polymorphism has also been found in the *GSTT1* gene, where around 20% of Caucasians and up to 60% of Asians have a homozygous deletion. Similarly, the GSTM1 null genotype was common in Caucasians (47–57%), Asians (42–54%), and Africans (16–36%). In contrast, the *GSTT1* null genotype was uncommon in Caucasians (13–26%), but common in Asians (35–52%) (Garte et al. 2001; Suthar et al. 2018). Hemlata Singh et al. (2022) found variance among inter- and intra-Indian ethnic groups in Himachal Pradesh, India, while Sharma et al. (2014) found variation among Delhi residents.

According to an analysis of epidemiological data conducted in the USA, broccoli (Wang et al. 2004) and crucifer vegetables provide *GSTT1*-positive people more cancer protection than null mutants (Spitz et al. 2000). Persons with GSTM1-positive status who ingested one serving of broccoli per week had a significantly lower prevalence of colorectal adenomas than those who did not eat broccoli. However, bigger intakes of broccoli may minimize the risk of adenomas in people. GSTM1 and GSTT1 nulls in Asian cultures may have stronger resistance to Chinese cabbage than GSTM1 and GSTT1 positives, according to research that used dietary frequency questionnaires (Zhao et al. 2001; Fowke et al. 2003) or detected ITCs in urine (Gasper et al. 2005). In those with a functioning *GSTT1* allele, Cornelis et al. (2007) found that eating *Brassica* plants was related with a decreased risk of heart infection. ITC metabolism or excretion is slowed down by having a GSTM1 or GSTT null genotype because GSTs are in charge of conjugating ITCs for excretion. Mercapturic acid is responsible for the metabolism of sulforaphane to a greater degree in persons with a deficiency in GSTM1 than in those with an abundance of GSTM1 (Steck et al. 2007; London et al. 2000). Because plasma has a low glutathione environment, GSTM1 activity may cause ITC and glutathione to be conjugated after being exported from epithelial cells (Traka 2016). SNP analysis was utilized by Armah et al. (2013) to examine genome-wide associations and identify an unique interaction between diet and genes. Participants who consumed glucoraphanin-rich broccoli in this study had their metabolic profiles affected by polymorphisms in the poly-A polymerase gamma (*PAPOLG*) gene.

There is a clear connection between the health advantages of increasing the amount of GSLs in one's diet and the findings of experimental studies on the relevance of GSL-rich diets in the prevention of sickness. GSL-rich diets improve carcinogen detoxification, affect critical metabolic processes, and alter signaling pathways that induce sickness. GSL-rich broccoli sprouts significantly increased carcinogen excretion in test participants in Qidong (China), a place with a high incidence of air pollution, especially in GSTT1-positive persons, implying that these individuals' phase 2 detoxifying potential was enhanced (Kensler et al. 2012; Egner et al. 2014). It has been shown that consuming broccoli that is rich in glucoraphanin may reduce bad cholesterol in those who are at risk for heart disease. Following a diet rich in glucoraphanin for a period of one year has a considerable effect on the fundamental transcriptional mechanisms driving carcinogenesis, such as the insulin signaling molecules and the transforming growth factor (TGF) transcription factors. (Traka et al. 2008). Sulforaphane was shown to be a dose-dependent inducer of selective cytotoxicity in *HeLa* cells by Sharma et al. (2011). Because of this, the activation of the genes *IL-1, COX-2,* and *Bcl-2*

that are associated with apoptosis and inflammatory processes may be drastically reduced. Kim (2021) investigated the impact of sulforaphane in the treatment of Alzheimer's disease utilizing preclinical indicators of amyloid, inflammation, and oxidative stress. Sulforaphane was also investigated by Kaiser et al. (2021) as a phytochemical with potential cancer-preventive effects.

3 Genetic Resources for Health-Related (HR) Genes

3.1 Carotenoids

Brassica vegetables, except white cauliflower, are a strong source of β-carotene. Kopsell et al. (2004) observed that genotype variation for β-carotene varied from 3.8–10.0 mg/100 g. For either lutein or β-carotene, there was no significant genotype-by-season interaction. It is also abundant in broccoli, where levels varied between 0.4 and 2.4 mg/100 g FW (Kurilich et al. 1999). In addition, a little amount of a-carotene was identified, although <0.1 mg/100 g FW. β-carotene is abundant in orange cauliflower, which is the most intriguing and alluring Cole vegetable (Crisp et al. 1975; Dickson et al. 1988). This feature was seen in cauliflower groups with snowball (or late) development, hence its initial utilization in other cauliflower ecotypes was limited. Cauliflower is a very thermosensitive crop due to its curdling behavior, which restricts the economic viability of the trait to a narrow age group. To maximize the potential of this natural mutant, it is necessary to introgress the trait into genotypes of various age groups. Through the process of crossing orange-curded *OrOr* plants from the heterozygous line with orange-curding genotype line 6932, Dickson et al. (1988) were able to develop three orange high-carotene cauliflower inbreds. These inbreds were given the names NY 163, NY 165, and NY 156. The NY 165 and NY 163 cultivars were derived from distinct F_2 plants from the same cross, whereas the NY 156 cultivar originated from a cross between an *OrOr* plant and a white cauliflower, NY 9120 GL.

Kalia et al. (2020) launched a drive to increase carotene content in popular Indian cauliflower cultivars by introducing a spontaneously dominant *Or* mutant gene. They hybridized *Or* lines in Indian cauliflower using the donor source of the *Or* gene, EC625883, as well as Pusa Meghna, CC14, and DC 41-5 from the early maturity group and Pusa Sharad, CC35, and DC 18-19 from the mid-maturity group.

Kalia et al. (2020) examined Indian cauliflower containing *Or* gene in advanced lines. The donor source of the *Or* gene, EC625883, as well as the recipient parents Pusa Meghna, CC14 and DC 41-5 from the early maturity group and Pusa Sharad, CC35 and DC 18-19 from the mid-maturity group participated in a hybridization procedure for this purpose. Figure 17 depicts the β-carotene concentration of top plants at BC_2F_1 with various genetic backgrounds in the early and intermediate maturity groups. In the background of Pusa Meghna (early maturity group), Pusa Sharad (DC309), and CC-35 (mid-maturity group), the β-carotene content of individual plants in the backcross generation BC2F1 ranged from 4–9 ppm. However, in the background of DC41-5 (early maturity group) and DC18-19 (mid-maturity group), the

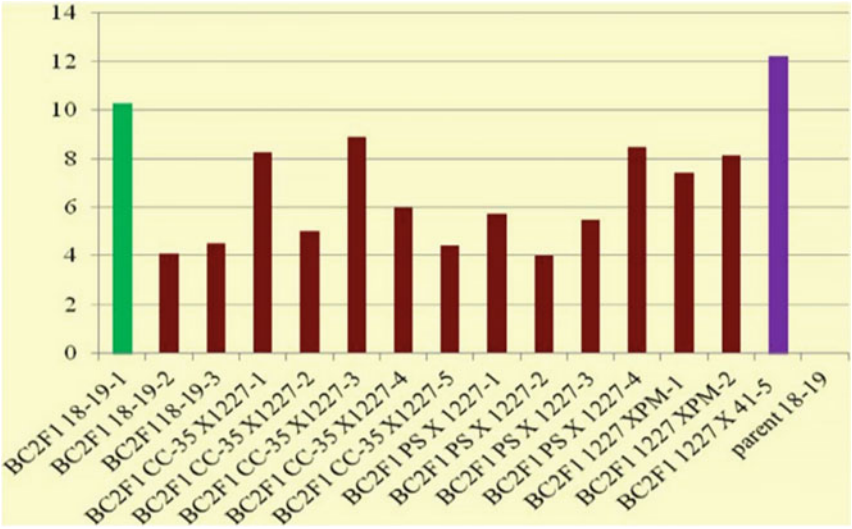


Fig. 17 BC_2F_1 elite selects for β-carotene content (ppm) in different genetic backgrounds of early (PM, 41-5) and mid-maturity (PS, CC35, 18-19) group with donor 1227 (EC625883). (Adapted from Kalia et al. 2020)

highest level was 12 and 10 ppm, respectively, when compared with the white wild parent. According to their findings, the β-carotene concentration of backcross progeny varied widely. The range of β-carotene concentration in the dark-orange category was DC 309 × EC625883 (2.03–14.64 g/g), CC-35 × EC625883 (7.03–15.99 g/g), and DC 18–19 × EC625883 (4.52–18.60 g/g) (Kalia et al. 2018). They reported combination of high recurrent parent genome in two plants with high β-carotene content, namely 18-19-1-7-7 (18.60 g/g) and 18-19-1-3-6 (17.94 g/g). BC1F1 was found to have more than 10 g/g β-carotene in these promising advanced lines.

In nine broccoli genotypes, lutein is the most abundant carotenoid, with concentrations ranging from 65.3–139.6% g/g dry matter (DM) of edible part. Violaxanthin was also observed in higher concentrations (17.9–35.4% g/g DM) than lutein. β-carotene and neoxanthin were found in concentrations comparable to violaxanthin (Farnham and Kopsell 2009). Since lutein is also crucial for human health, a decent breeding program should pay enough attention to lutein. Kopsell et al. (2004) identified *Brassica oleracea* L. var. *acephala* as the greatest source of lutein and β-carotene among *Brassica* vegetables. They investigated 23 cultigens and found that lutein concentrations varied from 4.84 mg/100 g fresh weight (343-93G1) to 13.43 mg per 100 g fresh weight (Toscano), while β-carotene accumulations ranged from 3.82 mg (30343-93G1) to 10.00 mg/100 g fresh weight (FW) (Toscano).

Kopsell et al. (2007) studied the carotenoids pigments of all six genetically related members of U's triangle. *Brassica rapa* has the greatest concentrations of antheraxanthin (0.79 mg/100 g fresh weight), lutein (8.89 mg/100 g fresh weight), and zeaxanthin (0.75 mg/100 g fresh weight). *B. juncea* had the greatest levels of total chlorophyll, and β-carotene at 125.9 and 4.41 FW mg/100 g, respectively.

Significant carotenoids in *Brassica nigra* were 5, 6-epoxylutein (0.41 mg/100 g FW), and violaxanthin (2.28 mg/100 g FW). The greatest concentration of neoxanthin (2.10 mg/100 g FW) was found in *B. oleracea. B. carinata* and *B. napus*, among the amphidiploids, collected much fewer carotenoids than the diploid species and *B. juncea*. *Brassica* plants provide distinctive health benefits when ingested. The data will be valuable for breeding programs aiming to boost carotenoids in *Brassica*.

3.2 Anthocyanin

Anthocyanin content in *Brassica oleracea* is due to natural mutant and probably transferred through natural crossing between these purple mutants and traditional color types (Volden et al. 2008). Because all the members of the *Brassica oleracea* are easily crossable and found in same localities (Piccaglia et al. 2002). In red cabbage, anthocyanins were in the range of 1.38 mg/g FW in red cabbage originated from the UK, 1.25 mg/g FW in Italian red cabbage, 1.14 mg/g in Norwegian fresh red cabbage, and 0.14–0.91 mg Cy/g fm in five tested varieties (Wiczkowski et al. 2014). Scalzo et al. (2008) detected a lower amount of anyhocyanins (0.77 mg/g FW) in red cabbage from Northern Italy (0.77 mg/g FW). The greatest concentration of these red pigments (3.22 mg/g FW) was found in American-collected red cabbage (Wu et al. 2006; McDougall et al. 2007). Strauch et al. (2019) analyzed 29 red cabbage genotypes for anthocyanin content and found a range of 11.3–30.1 mg/g, but Ahmadiani et al. (2014) reported 8–14 mg/g DW.

3.3 Glucosinolates

Wang et al. (2012) conducted research on five prominent Chinese cultivars and reported that the amounts of glucoraphanin in cultivated varieties ranged from 1.57–5.95 mol/g, but the levels in inbred lines ranged from 0.06–24.17 mol/g. The distribution of goitrin concentration in cultivars is about 1.77–6.07 mol/g with an average of 3.20 mol/g. Individual glucosinolates concentrations and the distribution of glucosinolate classes vary significantly among populations of broccoli. Wang et al. (2019) studied glucosinolates in the seeds of 32 pure lines and six commercial broccoli cultivars and discovered that aliphatic glucosinolates accounted for more than 90% of total glucosinolates. Glucoraphanin and glucoerucin were abundant in 27 samples, whereas progoitrin was prominent in seven samples. 4-(methylsulfinyl)butyl isothiocyanate and 5-(methylsulfinyl)pentanenitrile concentrations varied from 2.6–91.1 mol/g fresh weight and 0–35.4 mol/g fresh weight, respectively. Klopsch et al. (2018) discovered 23 GLS breakdown products in 91 genotypes of *Brassica rapa,* including nine isothiocyanates, ten nitriles, and four epithionitriles. Epithionitriles were the most common breakdown products because to the abundance of alkenyl GLSs. The researchers Sun et al. (2011) found that the levels of glucosinolate in the sprouts, rosette leaves, and bolting stems of each of the 27 different types of Chinese kale were distinct from one another. The overall and individual glucosinolate amounts of the

different plant species' edible portions varied. According to Cartea et al. (2008b), the total glucosinolate content ranged from 11.0–53 mol/g dw in kale, but from 10.9–27 mol/g dw in cabbage. They also discovered a temporal difference in total glucosinolate concentration, which ranged from 22 m/g dw during spring planting to 13 m/g dw during autumn sowing. Li et al. (2021) looked at eight developmental organs from 80 broccoli genotypes and found 12 GSLs in florets with concentrations ranging from 0.467–57.156 mol/g dry weight (DW). Lee et al. (2014) measured total GSL concentrations in 62 Chinese cabbage varieties ranging from 2.83–48.53 mol/g DW. Bhandari et al. (2015) investigated nine different *Brassica* varieties, including cabbage, cauliflower, broccoli, cabbage radish, pak choi, Chinese leaf mustard, kale, and baemuchae. Broccoli seeds and sprouts had the most total GSL (110.76 mol/g and 162.19 mol/g, respectively), whereas leaf mustard shoots and roots contained the least (61.76 mol/g and 73.61 mol/g, respectively). The total glucosinolate content of 36 *Brassica napus pabularia* group cultivars ranged from 1.4–41.0 mol g1 dw at one location and from 1.2–7.6 mol g1 dw at the other (Cartea et al. 2008b). Investigation of glucosinolate content of 65 *Brassica oleracea* accessions included broccoli (50), brussels sprouts (4), cabbage (6), cauliflower (3), and kale (3). The primary glucosinolates found in broccoli were glucoraphanin, gluconapin, and glucobrassicin. The content of glucoraphanin in broccoli varies between 0.8–21.7 mol/g dry weights depending on the variety. Sinigrin (8.9, 7.8, 9.3, and 10.4 mol g-1 DW) and glucobrassicin were the most abundant glucosinolates in brussels sprouts, cabbage, cauliflower, and kale (3.2, 0.9, 1.3, and 1.2 mol g-1 DW). Furthermore, brussels sprouts contained high amounts of gluconapin (6.9 mol/g DW). Branca et al. (2002) showed that Sicilian cauliflower types with colored curds contained higher levels of glucosinolates, namely glucoraphanin, than commercial cultivars with white curds. Carlson et al. (1987b) investigated the glucosinolates profile of *Brassica* vegetables, which ranged from 13.89–16.39 mol/g DW in broccoli, 13.0–14.17 mol/g fresh weights in brussels sprouts, 10.28–17.20 mol/g fresh weight in cauliflower, 14.61–21.72 mol/g FW in collard, 15.61–21.72 mol/g FW in kale, 18.50 μmol/g FW in mustard green, and 19.39 μmol/g FW in kohlrabi.

The GSL profiles of the screened wild Brassicas differed, with *B. macrocarpa* having a relatively high level of sinigrin and *B. rupestris* presenting a profile comparable to the crops investigated by Branca et al. (2002). *B. macrocarpa, B. villosa* var. *drepanensis*, and *B. atlantica* leaves were discovered to have high amounts of total GSL (11.88 M/g fresh weight) and sinigrin. Among the wild species, *B. macrocarpa* exhibited the highest GSL content (10.20 M/g FW leaves). The GSL concentration of *B. rapa* accessions varied from 1.49–3.06 M/g fresh weight (FW) leaves, whereas *B. rupestris* accessions had a GSL content of 1.71 M/g FW leaves. Leafy *B. rapa* subsp. *oleifera* is being used to breed *B. napus* oilseed cultivars with low glucosinolate concentration. Glucosinolates in wild *Brassica* species such as *B. montana* (2.17 mol/g fresh weight), *B. incana* (4.37–11.27 mol/g fresh weight), *B. rupestris* (11.88 mol/g fresh weight), *B. depranensis* (4.73 mol/g fresh weight), and *B. macrocarpa* (3.76 mol/g FW). 3-methylsulfinylpropyl, 3-methylthiopropyl, and 2-propenyl were the main glucosinolates in *B. villosa, B. drepanensis,* and *B. atlantica*, respectively are presented in Table 1. Differences

Table 10 Wild sources for nutritional traits in *Brassica* group

Nutrition	Sources	References
Tocopherols (Vitamen E)	91 crucifer spp.	
	*Diplotaxis viminea (*68)	
	Schivereckia doerfleri (2479)	
Flavonoids	*Erucastrum* spp.	
	Diplotaxis 13 spp.	
	Diplotaxis acris, D. harra	
	Camelina sativa	
	Lepidium sativum	
	Diplotaxis tenuifolia	
	Brassica spp.	
	Erucastrum spp.	
	Crambe spp.	
	Brassica spp.	
	23 taxa in tribe (sinapine)	
	51 crucifer spp.	
	Eruca vesicaria subsp. sativa	
	12 crucifer spp.	
	4 crucifer spp.	
	Brassica oleracea and 9 wild C genome spp	
	Cakile 13 spp.	
	Zilla 3 spp.	
	Camelina sativa	
Phenolics	*Brassica* spp. wild	
	259 crucifer spp.	
	9 crucifer spp.	
	Brassica 14 spp.	
	Brassica 20 spp. highest *B. montana*, *B. nigra* and *B. oleracea*	
	Lepidium peruvianum	Li et al. (2001)
	Family survey: 96 glucosinolates	Fahey et al. (2001)
	85 crucifer spp.	Bennett et al. (2004)
Glucosinolates	*Coincya* 3 spp.	
	Capsella bursa-pastoris	
	Eruca vesicaria subsp. *sativa*	
	Erysimum allionii	
	Lesquerella fendleri	
	Matthiola longipetala	
	Hesperis matronalis	
	Diplotaxis tenuifolia	
	Brassica 4 spp.	
	Lobularia maritime	

(continued)

Table 10 (continued)

Nutrition	Sources	References
	Kremeriella cordylocarpus	
	Plotaxis and *Eruca* spp.	
	Sinapis spp.	
	Erysimum cheiri	

Source: Wild Crucifer Species as Sources of Traits, https://www.brassica.info

in glucosinolate profiles were attributed to changes in alleles at the *GSL-OXID* and *GSL-ALK* loci (9). Faulkner et al. (1998) investigated the glucosinolate profiles of three wild *Brassica* species while introducing beneficial glucosinolates into chromosomal areas. They discovered that the most prevalent glucosinolate in *B. villosa* was 3-methylsulfinylpropyl (glucoiberin) (119.0 moles/g dry weight), followed by 3-methylthiopropyl (glucoiberverin). These substantial glucosinolates account for about 95.9%, 82.4%, and 97.4% of total glucosinolates, respectively.

Several studies have assessed alien species of *Brassica* crops and identified certain of them as potential donors of quality traits (Table 10).

4 Classical Genetics and Traditional Breeding in Vegetable Brassicas

4.1 Genetics of HR Genes: Inheritance and Mode of Action

4.1.1 Genetics of Carotenoids Accumulation in Cauliflower Exploitation of *Or* Gene in Indian Cauliflower Improvement

Crisp et al. (1975) selfed the plants of clones of orange mutant which segregated and the following observations were reported: (i) three plant phenotypes were having white curds, plants with orange curds resembled parent plants, and plants with very small loose curds have more intense orange color in the ratio of 1:2:1, (ii) a semidominant (incomplete) gene action for orange curd trait and (iii) orange color was present in all the tissues of the curd except the pith and the color was due to xanthophylls and carotene. The plants with *OrOr* genotype showed intense orange color, whereas *oror* types were white. The gene for orange curd phenotype was designated as "*Or*" (after the color of curds).

Cauliflower as Ideal Crop for β-Carotene Biofortification

Following points highlight the significance of cauliflower as an ideal crop for β-carotene biofortification to counter the global problem of β-carotene deficiency, particularly through food-based approach:

1. Cauliflower is one of the ideal crops for public nutrition because its annual production is 25.53 million tonnes, predominantly in region of vitamin A

deficiency (Fig. 9). It has high productivity both in terms of per unit area and time scales.
2. The *Or* gene is a native spontaneous mutant in cauliflower itself, so it is not a transgene. Thus, it eliminates any prohibition and restrictions (as in case of transgenics or genome-edited crops) for commercial release and consumer acceptance. Thus, the biofortified varieties will have immediate reach to the end users and can show the health benefits in target group.
3. The phenotype of *Or* gene expresses in meristematic tissues of curd part which constitute the maximum share of edible portion. The harvest index (curd weight/gross plant weight) ranges from 0.40–0.55 kg in fully developed curds which highlight high recovery of the orange curd part, thereby β-carotene from a unit area.
4. Cauliflower has seven ecotypes, namely Italians or originals, Cornish, Northerns, Roscoff, Angers, Erfurt and snowball, and Indian cauliflower which are grown across the world. Additionally, Australians were also separated out from European types. This widespread distribution and natural evolution of crop highlights the immediate cultivation of *Or*-biofortified local ecotypes.
5. It is very common and grown at large scale in developing world which has major chunk of vitamin A–deficit human population. Thus, biofortified cauliflower can serve in a better way to the vitamin A vulnerable population in this region, a malnourished population.
6. Globally, it has large consumer base and have better transportability which highlights its scope for covering farming and nonfarming population.
7. The cooking processes of cauliflower include oil which enhances bioavailability of β-carotene as it is a fat-soluble vitamin.
8. White cauliflower is considered as bleached type and contains only glucosinolates as major functional compound while introgression impart additional health compound (i.e., β-carotene).
9. Despite nutritional significance, the orange cauliflower represents only a fraction of the traditional white cauliflower market, and still confined to specific markets as a niche product.

4.1.2 Genetics of Anthocyanin in *Brassica* Vegetables

Cabbage Head Color

Anthocyanin pigmentation in red cabbage was the subject of several genetic researches, and it was shown that this feature is passed from a single gene (Yarnell 1956) to multiple genes (Kristofferson 1924; Dickson and Wallace 1986). The red color of cabbage is influenced by at least two additional factors in addition to an increase in anthocyanin content (Kristofferson 1924). A single locus has several alleles for two dominant color components that are complementary to one another in order to explain earlier findings. The interaction of the genes *A* and *D* results in the production of red cabbage. *A* third gene, *B*, which produces the dark violet midrib when paired with *A*, produces the brilliant red midrib. While *C* is colorless on its own, combining it with *A* results in a violet midrib. Although *C* has the same effect

as *B* alone, *E* intensifies the dark violet color that *B* produces. Therefore, green cabbage with a red midrib is *daBCE* whereas crimson cabbage is *DABce* (Magruder and Myers 1933). Clearly, these studies and others demonstrate the existence of the quantitative traits of color inheritance (Kwan 1934).

Cauliflower Curd Color

The four major curd colors of cauliflower are white, green, purple, and orange. In purple cauliflower 'Grafitti' and Sicilian purple 'PC-1', this phenoptype is governed by a partly single dominant allele (Chiu et al. 2010; Singh et al. 2020). Purple Sicilian cauliflowers, which are broccoli and cauliflower hybrids (Gray 1992), on the other hand, exhibit recessive purple when mated with white cauliflower. Three additional "bleaching genes" are thought to be responsible for the whiteness of sun-exposed white curd (Crisp and Tapsell 1992). A spontaneous mutation of a single semidominant gene, *Pr*, resulted in a mutant cauliflower with intense violet colors (purple gene).

The major flavonoids derivative in the mutant's various tissues was cyanidin-3-(coumaryl-caffeyl)-glucoside-5-(malonyl)-glucoside. Curd, premature seedlings, premature leaves, premature flower buds, siliques, and seed endosperm are all rich purple, but adult leaves, stems, flower petals, and siliques are all colorless. The *Pr* gene provides instructions for making a tissue-specific R2R3 MYB signaling pathway, which is associated with the abnormal aggregation profile of anthocyanin in the mutant. The curds contained around 3.75 mg g-1 cyanidin-diglucoside (Chiu et al. 2010). Singh et al. (2020) examined the genetics of the 'PC-1' genotype established by repeated segregation of Sicilian purple material. The $173F_2$ population was segregated, showing that a gene dominant over the white allele governed the curd's purple color, although only partly. This was shown by the fact that the purple coloration was only partially exhibited. The vast selection of anthocyanin concentration in F_2 (3.81–48.21 mg/100 g fresh weight) was substantiated by the fact that now the purple coloration of the curds varied from extremely faint to vivid. This demonstrated that the anthocyanin level varied widely. In the F_2 population, co-segregation was not seen. The transcription of the BoMYB1 gene was greater in the purple curd genotypes 'PC-1' and 'Graffiti' than in the 'DC-466' genotype, but the synthesis of the *BoMYB2* gene was more in the genotype 'PC-1' than in the genotype 'Graffiti'. The existence of "broccoli-type" F_2 individuals and the genetic longevity of these individuals in $F_{2:3}$ imply that 'Sicilian purple' is an interim form between cauliflower and broccoli (Calabrese). It seems that the PC-1 and Graffiti genes behave differently since there is no correlation between the color of the curd and the coloration on the apical stem and leaf sections.

Broccoli Head Color

Broccoli contains anthocyanin, a substance with important antioxidant characteristics that may boost health, increase life expectancy, and prevent illness (Chaudhary et al. 2018). The accumulation of anthocyanins in broccoli inflorescences, especially in septals, provides a green/blue to violet tint. 'Purple Sprouting Early' is bred for its high anthocyanin content, which results in strikingly purple flower heads.

Cauliflower, cabbage, broccoli, and possibly knol khol all exhibit a variation of the *MYB DOMAIN PROTEIN 2 (BoMYB2)* gene that contributes to the purple coloration of *Brassica oleracea* (Yan et al. 2018). However, green-purple broccoli cultivars are not as desirable as the fully green variety, particularly in the Chinese market (Yu et al. 2019). This green-purple cultivar is climate sensitive, with cold temperatures inducing and intensifying the purple tint. Yu et al. (2019) used a DH population and SLAF sequencing to identify the purple sepal feature; three QTLs were discovered, with qPH being the dominant locus. C01-2, which is located on LG1, as well as two qPH loci. C01-4 is used in conjunction with qPH. C01-5 is just close to qPH. C01–2 (Reshma et al. 2018).

Other *Brassica* Vegetables

Chinese cabbage and pak choi are the most popular types of the *Brassica rapa* vegetable genus in China and other East Asian countries (*Brassica rapa* L. ssp. *chinensis*). There are several red or purple *B. rapa* cultivars, including multiple varieties of pak choi, turnip, and zicaitai. Notable was the purple leaf trait of Chinese cabbage transmitted from a lineage of *Brassica juncea* (Sun et al. 2006). In the fast-cycling *Brassica rapa*, a nonpurple mutation was recessive (Burdzinski and Wendell 2007), and the purple allele anl was initially identified on the Chinese cabbage linkage group R09. On the A09 linkage group, a significant gene influencing the purple petiole phenotype of Chinese cabbage was identified, and partial dominance for the trait was shown in flowering stalk (*Brassica campestris* L. ssp. *chinensis* var. *perperea* Hort) (Zhang et al. 2008). Two significant QTLs, BrEGL3.1 and BrEGL3.2, were discovered on the A09 linkage group (Guo et al. 2015). Purple gene from pak choi (*Brassica campestris* L. ssp. *chinensis* [Lour.] var. *cammunis* Tsen et Lee) was governed by a single dominant allele (Zhang et al. 2011), and this gene was mapped on A03 linkage group (Liu et al. 2013) and then fine-mapped on 54.87 kb interval. Anthocyanins have been identified as *B. rapa's* purple pigments. A comparison of the completed genome of *A. thaliana* and *B. rapa* led to the identification of anthocyanin biosynthetic genes (ABGs). Anthocyanin production pathways including 73 genes have been identified in *B. rapa*. *BrABGs* were created in many copies by whole genome duplication (WGD) and retained synteny with their *A. thaliana* orthologs. However, gene loss during whole genome triplication (WGT) led to the majority of genes having less than three copies. The upstream structural genes of the anthocyanin biosynthesis pathway have been retained more than their downstream counterparts. On the basis of the existence of these comparable structural genes, the anthocyanin production pathway in *B. rapa* was later discovered. Comparing copy numbers suggests more negative regulatory genes than positive have survived (Guo et al. 2014).

4.1.3 Chlorophyll-Rich Green Cauliflower

Zhou et al. (2011b) showed that combining colored curd to cauliflower improves both its appearance and nutritional value. Cauliflower is green due to the aberrant improvement of chloroplasts in its curd tissue. A partially dominant gene, according to genetic theories, is determinant for the green color (Crisp and Angell 1985).

However, it remains unknown what the inherited biological basis seems to be. Tan et al. (2020) employed QTL-seq analysis to locate genome areas crucial for the cauliflower green curd feature and to restrict targeting intervals for putative specific genes. The F_2 population is the result of a hybridization between a Stovepipe and an ACX800 (white curd × green curd) cauliflower plant. Two quantitative trait loci (QTLs) were identified on chromosome group of 7 (Gr7.1) and 5 (Gr5.1) using SNP analysis and complete DNA resequencing of green and white F_2 bulks (Gr7.1). Using CAPS markers and traditional genetic mapping, Gr5.1 was identified as a large QTL, but Gr7.1 had a little effect. Utilizing additional CAPS markers, high-resolution mapping of Gr5.1 in the second large F_2 generation reduced the target region to 0.3 cM and 236 Kbp, respectively (Chi et al. 2012). There are 35 genes in this region, four of which are UMP kinase and DEAD-box RNA helicase (Schmid et al. 2019). This is owing to the discovery that other plants with UMP kinase mutants and DEAD-box RNA helicase mutants form chloroplasts and have pale leaves (Chen et al. 2018b).

4.1.4 Genetics of Glucosinolates in *Brassica* Vegetables

Genetics of Aliphatic Glucosinolates

Four different studies were conducted and published in the 1990s under the title Genetics of Aliphatic Glucosinolats (GSL) (GSL I, GSL II, GSL III, and GSL IV) (Giamoustaris and Mithen 1996). Within the scope of these studies, models of the primary loci that control aliphatic GSL synthesis in *Brassica* were developed (Parkin et al. 1994; Mithen et al. 1995). This study's first section examined how distinct recombinant populations of *B. napus* displayed distinct patterns of glucosinolate properties (Magrath et al. 1994). Pentyl and butyl GSLs are synthesized by alleles at two extra loci, while propyl GSLs are formed by the side chain extension of amino derivatives (GSL-ELONG-A and GSL-ELONG-C), which is determined by genes at a GSL-PRO locus. The same population of *Brassica napus* provided evidence that supported a hypothesis in which genes at a simple Mendelian control the occurrence or absence of propyl G. Parkin et al. (1994) explored at the hydroxylation of alkenyl GSLs and found two alleles implicated in this process in both leaves and seeds. The *GSL-OH-A* locus in LG 3 has a much less impact than the *GSL-OH-C* locus in LG 13. Mithen et al. (1995) established that chromosome 4 alleles are responsible for two distinct types of side chain modifications in *A. thaliana*. First, methylsulfinylalkyl GSLs are changed into alkenyl GSLs (GSL-ALK), and subsequently methylsulfinylpropyl GSLs are transformed into hydroxypropyl GSLs (GSL-AL) (GSL-OHP). In 2001, Li et al. designed a conceptual framework for GSLs with three to four carbons. The GSL-PRO allele results in three carbon-containing GSLs, while the *GSL-ELONG* allele results in four carbon-containing GSLs (4C) (3C). By sequencing a homolog of the *Arabidopsis* GSL-ELONG gene present in the *B. oleracea* plant in 2002, the presence of 4C GSL was determined (Li and Quiros 2002). Key GSL genes in *B. oleracea* were identified and sequenced by mapping a BAC from this plant to its analogous area in *Arabidopsis*. This comparison was conducted to identify the presence or absence of the genes (Gao et al. 2004). All three 2-oxoglutarate-dependent dioxygenase (AOP) genes

Table 11 Genes involved in GSL biosynthesis in *Brassica* vegetables

Loci	Function	Broccoli	Purple cauliflower	White cauliflower
BoGSL-PRO	Regulate the production of propyl GSL (3C)	–	–	+
BoGSL-ELONG	A gene for side chain elongation that causes 4C GSL	+	+	–
BoGSL-ALK	Gene plays a role in the desaturation of the GSL side chain	–	+	+

Adapted from Li and Quiros (2002)

in *Arabidopsis* are essential for the generation of GSL. These genes are designated *AOP1 (null allele), AOP (GS-ALK),* and *AOP3 (GS-OHP)*. There are many copies of *AOP2.2 (BoGSL-ALKb), AOP3, AOP1.2, AOP1.1,* and *AOP2.1 (BoGSL-ALKa)* in *B. oleracea*, however, only two of these genes are absent. Table 11 displays the evolution of genes responsible for the formation of glucosinolates in three closely associated plants: broccoli, Sicilian purple, and cauliflower.

Glucosinolate side chains are oxidized and hydroxylated by the BoGSL-ALK, *BoGSL-OXID*, and *BoGSL-OH* genes, respectively. On their own, BoGSL-ELONG and *BoGSL-PRO* separate. One of the most essential genes influencing the quantity of glucoraphanin in *Brassica* plants is the *A. thaliana AtAOP2* homolog, *BoGSL-ALK*, which produces a 2-oxoglutarate-dependent dioxygenase requisite for the transition of methylsulfinylalkyl GSLs to their alkenyl derivatives. Total GSLs are calculated when methionine enters the pathway controlled by MAM genes at *GLS-ELONG* loci. The allelic variation at the *GLS-OX, GLS-ALK,* and *GLS-OH* loci is then used to determine the GLS profiles. Profile selection is possible because *B. oleracea* generally exhibits a higher degree of variation at these loci than *B. rapa*.

The transformation of a methylthioalkyl GSL structure into a methylsulfinylalkyl structure is carried out by the GSL-OX gene (Mithen et al. 1995). Numerous modifier loci were detected during a forward genetics screen of *Arabidopsis* mutants, which complicated genetic attempts to pinpoint the source of this characteristic (Kliebenstein 2009). Hansen et al. (2007) identified a considerable variety of possible key genes that co-expressed with genes previously known to be involved in the aliphatic GSL production pathway using accurate variation mapping. It was discovered that a group of five flavin monooxygenases peculiar to vegetable Brassicas governs the transformation of methylthioalkyl to methylsulfinylalkyl GSL. Through the use of *Arabidopsis* sequence analysis, Chinese cabbage, an EST library, a *B. rapa* cultivar, and BAC microarrays, Zang et al. (2009) were able to determine the genes involved in GSL synthesis and regulation. Apart from *CYP79F2, FMOGS-OX2,* and *AOP3,* they discovered 44 *Arabidopsis* homologues. These genes are available in a variety of copies. Although the numbers of genes had evolved between the two species, they found considerable intercorrelations in the GSL. This highlights the development of the link between *B. rapa* and *Arabidopsis* (Zang et al. 2009). The *B. rapa* genome sequence was then put to use in a more thorough investigation (Wang et al. 2011). Every one of the 102 GSL-producing

alleles in *B. rapa* were linked to a single chromosome, as determined by the scientists. The majority of the duplicated *Brassica* GSL genes share synteny with those in *B. rapa* and *Arabidopsis* (93%). Additionally, a triplication event in *B. rapa* is associated with variation in copy number in these genes, and GSL gene content may account for both GSL profiles and accumulation in *B. rapa* (Wang et al. 2011). Comparative genomics using the *Arabidopsis* genome sequence was used to find the bulk of *Brassica* GSL gene sequences. In recent years, several studies have examined the significance of the most important GSL genes in *Brassica* crops. From the leaves of pak choi (*B. rapa*), the genes *CYP83A1* and *CYP83B1* were synthesized, and their expression in various tissues was analyzed. These genes catalyze oximes to produce aliphatic, aromatic, or indole GSLs. There was a link between the expression patterns of these genes and the aggregation of GSL in many genotypes and cells of this species (Zhu et al. 2012). Another study used homologous recombination to genetically modify aliphatic GSL backcross lines (Hirani et al. 2013). In each backcross generation, a resynthesized *B. napus* line was crossed with a Chinese cabbage–doubled haploid line, and nonfunctional genes were discovered using marker-assisted selection (BC_3F_2). Recurrent parent progenies bearing the *GSL-ELONG* gene showed a decrease in 5C GSLs, viz. glucoalyssin, gluconapoleiferin, and glucobrassicanapin (aliphatic GSL). The experimental analysis showed that the functional allele was switched for the nonfunctional *B. oleracea GSL-ELONG* allele. Augustine et al. (2013) revealed that the chromosomes of *B. juncea* (AABB), *B. nigra* (BB), and *B. rapa* (AA) each have four *MYB28* homologues. Four genes are responsible for encoding the functional *MYB28* proteins, which results in the production of aliphatic GSL with a comparable structure and composition. Subsequently, they found four CYP83A1 homologues and utilized functional studies to demonstrate that each of these four CYP83A1 homologues preserved extensive, interconnected, but different transcriptional activity across a vast array of *B. juncea* cell and tissue types. This was accomplished by demonstrating that each of those four CYP83A1 homologues sustained wide-ranging, interrelated, but distinct expression patterns (Meenu et al. 2015). In a second study, quantitative real-time PCR was utilized to analyze the transcription of multiple aliphatic GSL regulating sequences, including the *BrMYB29* and *BrMYB28* genes, across different tissues and morphogenesis of *B. rapa* ssp. *pekinensis* (qRT-PCR). An overlapping gene regulation pattern was noticed between BrMYBs and the genes downstream from them at different stages of embryonic development (DSGs). At most stages of embryonic development, the *BrMYB29.1* and *BrMYB28.3* genes were present in greater numbers than the other *BrMYB* paralogs. These genes comprised paralogs of the *BrMYB29* and *BrMYB28* genes. This was true for all of the *BrMYB* paralogs (Baskar and Park 2015). The function of the GSL genes in *B. oleracea* was not understood until relatively recently. A comparative analysis of databases, Ensemble and Bolbase, identified that 84 genes in *B. oleracea* are orthologous to those in *B. rapa* and are involved in GSL production, regulation of gene expression, and degradation. It was determined that these genes are orthologous to one another. Yi et al. (2015) examined the levels of gene expression in a total of 12 unique *B. oleracea* genotypes that belonged to four distinct groups. It has been shown

that each of these genes may manifest itself in a distinct tissue type. Only eight of them came out as cauliflower florets that were produced in kohlrabi stems; the rest of them came out as cabbage or kale leaves.

Indole Glucosinolates

Brassicaceae plants employ indolic GSLs, a kind of secondary metabolite generated from tryptophan, as a chemical weapon in their arsenal. *Arabidopsis* has contributed as a paradigm for dramatic improvements in indolic GSL synthesis, transport, and performance characteristics in recent years (Naur et al. 2003). *Brassica* biogenesis genes were found in *B. oleracea* (Gao et al. 2014), *B. rapa* (Yun-Xiang et al. 2008; Frerigmann and Gigolashvili 2014), and *B. napus* (Pfalz et al. 2011; Hirschmann and Papenbrock 2015; Klein et al. 2006) by using EST sequencing, BAC libraries, whole-genome sequence information, and metabolic profiling. Tryptophan is transformed into indole-3-acetaldoxime, an aldoxime, as a precursor amino acid by the *CYP79B2* or *CYP79B3* gene products (Wittstock and Halkier 2000). The enzymes *CYP83A1* and *CYP83B* transform aldoximes into activated chemicals (either acid nitro compounds or nitrile oxides). With the use of metabolic engineering on *B. rapa*, scientists investigated the role of this CYP gene family that is thought to be involved in the early stages of the process. With cDNA expressing *Arabidopsis CYP79B2/CYP79B3* and *CYP83B1* Chinese cabbage plants were altered. They found that activation of *CYP79B3* or *CYP83B1* had no effect on the accumulation of indole GSL by analyzing the total GSL of different acessions transformed with each triple and double construct. When *CYP83B1* was stimulated alongside *CYP79B3* and/or *CYP79B*2, methoxyglucobrassicin hydroxyglucobrassicin and glucobrassicin levels increased. The expression of two more enzymes must be coordinated in order to change *B. rapa*'s production flow toward indole GSL. Additionally, overexpression of the *CYP79B2* or *CYP79B3* genes results in a rise in indole-3-acetic acid (IAA), indicating that these genes are crucial for the production of both auxin and indole GSL. When *BrCYP83B1* was identified and described in the pak choi leaves (*B. rapa*), researchers looked at the sequence of its transcriptional activation in a variety of pak choi genotypes and tissues (Zhu et al. 2012). Recent transcriptome experiment by Gao et al. (2014) identified the CYP family of genes (*BoCYP79B2/BoCYP79B3/BoCYP83B1*) in *B. oleracea*. These important GSL synthesis genes were found to be substantially conserved across the Brassicaceae family by the researchers using amino acid sequence analysis and sequence alignment. Thiohydroximic acids are converted to desulfoGSLs by the glucosyltransferases *UGT74B1* and *UGT74C1* of the *UGT74* family (Grubb et al. 2004). Indolyl methyl-thiohydroximate is changed into indolyl methyl desulfo GSL by the enzyme UGT74B1 along the indolic GSL biogenesis. When BnUGT74B1 upregulation in *Brassica napus* strains elevated aliphatic and indolic GSL levels, the severity of disease symptoms and tissue damage in reaction to an infection with *Sclerotinia sclerotiorum* and *Botrytis cinerea* was reduced. This was seen in both cases (Zhang et al. 2015). However, the homologues of this gene that are found in *B. oleracea* were not present in the BAC or cDNA libraries of *B. rapa* (Gao et al. 2014). Sulfotransferases (SOTs) accelerate the last stage of GSL production, which is a crucial part of GSL biosynthesis. There are three SOTs in *Arabidopsis*, SOT16,

SOT17, and SOT18. Despite the fact that the three enzymes' affinities vary, all three employ desulfoGSLs as substrates (Klein et al. 2006). In this specific model plant, the substrate indolic desulfoGSLs is SOT16's first preference (Klein et al. 2006). The recent release of the full sequencing of *B. napus* enables it to conduct an in-depth study of the SOT family of proteins in its entirety (Chalhoub et al. 2014). Four *Arabidopsis* SOT16 homologues were discovered in the *B. napus* genome by Hirschmann and Papenbrock (2015) using sequencing. Studies of the substrate preferences of BnSOT16 *in vitro* revealed similarities to *Arabidopsis* orthologs. On the side chains of indole GSLs, hydroxylations and methoxylations are catalyzed by a number of enzymes. An unidentified methyltransferase transforms glucobrassicin from *CYP81F2* to hydroxyglucobrassicin, which is subsequently converted to metoxyglucobrassicin (Pfalz et al. 2009). The significance of two CYP81F4 isoforms found in the *Brassica rapa* genome was determined through heterologous complementarity and functional studies of the matching *Arabidopsis* variation (Wiesner et al. 2014). Additionally, a *B. oleracea* transcriptome investigation discovered the existence of the genes *BoCYP81F4, BoCYP81F1*, and *BoCYP81F3* (Gao et al. 2014). *MYB34, MYB51,* and *MYB 122* are part of a complex network of TFs that control the synthesis of indolic GSL in *Arabidopsis* (Hirschmann and Papenbrock). *MYB51* largely controls indolic GSL production and is associated with biotic stress responses, but *MYB34* and *MYB122* also promote auxin biosynthesis (Celenza et al. 2005; Gigolashvili et al. 2009). Using *BLAST N* and *BLAST P* on the drought *B. rapa* genome and annotated genes, Wang et al. (2011) revealed comparable *Arabidopsis* MYBs linked to indolic GSLs. Each transcription factor from *B. rapa* shared more than 70% of the sequence with transcription factors from *Arabidopsis*. Researchers found *BoMYB51* in broccoli seeds (Araki et al. 2013) and sprouts and *BoMYB34* in kale using RNA-seq analysis and EST data (Gao et al. 2014). In addition, in response to environmental signals, TFs may alter plant metabolism. Using signaling molecules and mechanical stress, GSL levels have been raised in a number of *Brassica* species. There is evidence from several studies on *Arabidopsis*, *Brassica oleracea* (broccoli) (Koritsas et al. 1991; Bodnaryk 1992), *Raphanus sativus* (Red radish [Doughty et al. 1995] and China Radish rose [Brader et al. 2001]), *Brassica rapa* (turnip), and *Brassica napus* (rutabaga, cabbage, and oilseed rape) that methyl jasmonate and wounding may affect indolic GSL levels (Baenas et al. 2014). Hormone therapy enhances TFs in *B. rapa* MYBs encoding indolic GSLs, per prior *Arabidopsis* findings (Frerigmann and Gigolashvili 2014). The experimental analysis of Chinese cabbage seedlings with jasmonate, salicylic acid, ethanol, and abscisic acid increases *MYB34* and *MYB122* expression, which alters the expression levels of SOT16 and CYP79B3/CYP79B2 resulting in an increase in the accumulation of indolic GSLs (Thiruvengadam et al. 2015). As a result, the types and quantities of indolic GSL in diverse plant organs are greatly influenced by environmental conditions.

Modifying GSL Content in Vegetable Brassicas

In recent times, there has emerged a considerable amount of fascination with the creation of procedures and techniques that can modify the levels of specific GSLs in *Brassica* plants. This interest can be attributed to the fact that these techniques and

procedures have the potential to improve crop yields. This is due to the fact that although certain GSLs have desired qualities such as flavors, biofumigation, resistance to insects, and anticarcinogenic, others have unfavorable qualities such as harshness and goiter disease. In addition, some GSLs have desirable properties such as preventing cancer (van Doorn et al. 1998). One goal of improving *Brassica* is to increase the amount of positive GSLs and decrease the number of negative GSLs (Traka and Mithen 2009). Either this is done in an effort to grow crops with a high value and improve the overall quality of food, or it is done in an effort to collect data that can be utilized to investigate how these compounds influence living things (Fahey et al. 2001; Cartea et al. 2008a). In addition to structural variety, the genetic similarity level (GSL) diversity that exists across families, genera, species, subspecies, and subspecies accessions is considerable (Cartea et al. 2008b; Kushad Mosbah et al. 1999; Bradshaw and Wilson 2012). Because of this variety, it is now possible to design new sorts that have GSL content and composition that is acceptable. It has been shown to be possible to adjust the GSL content of cruciferous vegetables via a variety of techniques, including as classical breeding techniques and genetic manipulation. This modification has been shown to be practical. In recent years, this goal has been accomplished.

4.2 Breeding Objectives: Positive and Negative Selection

Higher crop yield, adaptation to wide climatic conditions, and resistance to common diseases and pest are major aims of breeding *Brassica* vegetables. In quality traits, higher levels of β-carotene and anthocyanin in edible portion of *Brassica* vegetables are attracting growers and consumers. Higher levels of beneficial glucosinolates (glucoraphanin and glucoiberin) and lower levels or complete elimination of progoitrin are interesting areas for quality breeding in *Brassica* vegetables. In addition to these, higher levels of other antioxidants such as ascorbic acid, lutein, dietary minerals and vitamin A are also targeted by the quality breeders for public health. Increase in chlorophylls is required for better consumer acceptance. It is also essential to increase the sugar fraction in *Brassica* vegetables so that the taste can be further improved, particularly, in higher temperature conditions when bitterness detracts many consumers.

4.3 Classical Breeding Achievements: Composition and Contents

4.3.1 Breeding for Carotenoids

To obtain 100% heterozygous plants (*Oror*), the following two ways can be adopted: (i) use combination of white homozygous recessive (*oror*) SI/CMS line × homozygous dominant (*OrOr*) as male or (ii) dominant homozygous (*OrOr*) orange as SI/CMS line × homozygous recessive (*oror*) as male. For this, the *Or* gene need to be transferred to a self-incompatible/CMS line of a respective maturity groups. Since, originally the *Or* genotypes were found in snowball group, and now since it

has been introgressed in all the maturity groups of Indian cauliflower in an attempt to develop β-carotene-rich cultivars to tackle prevalent malnutrition in Indian population, its utilization in hybrid development needs three-step activities. First, introgression of *Or* gene into a fertile non-SI line of the target maturity group; second, convert the *Or* line into a CMS lines; and third, explore/utilize the homozygous orange CMS lines in hybrid development. Alternatively, the fertile non-SI *OrOr* lines developed after first phase can be used as male parent along with white curding (*oror*) CMS lines of respective maturity groups to develop hybrids. The use of SI^{r} or CMS^{r} as one of the parents in hybrid breeding is essential because hybrid seed production in cauliflower is entirely dependent upon these genetic mechanisms.

4.3.2 Development of Open-Pollinated Varieties

Crisp et al. (1975) suggested to exploit the *Or* gene using hybrid breeding only because the *Or* gene in homozygous state confront with curd phenotype reducing their market value. Since the orange curd phenotype is governed by a single semidominant (partial dominant) *Or* gene, it has penalties on different morphological, developmental, phonological, and yield traits. Thus, the only plants heterozygous for *Or* gene have commercial value and the hybrid seed production is abundant. During seed production, the occurrence of homozygous plants (*OrOr*/*oror*) in open-pollinated varieties (OPV) will give undesirable phenotypes, particularly homozygous orange (*OrOr*) plants. The white curding plants are removed at curding stage and marketed from the seed production block to avoid contamination. These homozygous orange plants rarely produce satisfactory seeds. The curds of these plants remain very small (30–40 mm or less in diameter; <80 g in comparison to normal curd of 150–200 mm), deformed, and unmarketable (Dickson et al. 1988; Singh and Kalia 2021). The homozygous recessive (*oror*) plants produce white curds which are normal in shape and of full size as per the parental line, hence they have proper market for them as white curds. Thus, only heterozygous orange color plants are advanced for seed production in case of open-pollinated varieties. In the next year, this OP variety seed lot produces plants of three different types as white full size curds: intermediate orange full size curds: intense orange very small size curds in a ratio of 25:50:25, respectively. Thus, from such varieties, almost 75% plants are marketable, of them 50% are orange and 25% are white.

4.3.3 Development of Orange Hybrids in Cauliflower

Four F_1 hybrids by using white self-incompatible (SI) F_1 (90, 91, 95, and 96) of two isogenic lines differing in incompatibility alleles and a high carotene orange curd line as male parent. The female lines were four early white SI hybrids of two isogenic lines differing by incompatibility alleles. The orange curd hybrids so developed showed a variation in β-carotene content which ranged from 70.6–293.4 μg/100 g fresh curd tissues. The orange hybrid 95 × orange inbred line and 96 × orange inbred line had maximum β-carotene content (293.4 μg/100 g fw), which was almost 14.7 times higher than the white curd female parent, but around 3.6 times less than the orange male parent (1060 μg/100 g fw). Out of four orange hybrids, two were from female parental F_1s 90 and 91 (isogenic SI hybrids). These had lesser β-carotene content of

Fig. 18 CMS-based promising orange F_1 hybrids in early group of Indian cauliflower. (Photo: Shrawan Singh)

139.7 and 70.6 μg/100 g fw, respectively, than another set of orange hybrids 95 × orange inbred line (293.4 μg/100 g fw) and 96 × orange inbred line (155.2 μg/100 g fw). The author's group also observed strong influence of female parent on intensity of orange color and β-carotene content in Indian cauliflower.

Ding and Jian (2010) developed an orange-colored cauliflower hybrid 'Jinyu 60' using a white curding cytoplasmic male sterility (CMS) CMS92-60 as female parent and an orange-colored inbred line 93-4 as male parent. The F_1 hybrid was superior to the parental lines for both nutritional quality and commercial traits. The hybrid 'Jinyu 60' produce orange curds and the β-carotene content was 31.3 μg/g in dry curds, which was almost 10.8 times higher than that of the white curd. However, the vitamin C, dietary fibre, protein, and mineral elements were nearly equivalent to those of white cauliflower parental line. The curd weight was reported to be in the range of 800–1000 g and maturity period is about 60 days after transplanting.

The author's group at ICAR-IARI, New Delhi, India, also developed orange-colored cauliflower hybrids, namely *DCEH-15419* and *DCEH 529819*, using white curding CMS lines as female parent and intense orange-colored fertile inbred lines as male parents in tropical Indian cauliflower (Fig. 18). Both hybrids produce all orange curds at a higher mean temperature range (20–25 °C), hence promising for growing in tropical regions. The yield potential of these promising hybrids is around 18 t/ha with standard crop management practices (IARI Annual Report, 2021). Several hybrids in market are available, particularly in European-type cauliflower; some of them worth mentioning are 'Cheddar', 'Orange Burst', and 'Orange Bouquet'.

4.3.4 Breeding for Anthocyanin

Recurrent Breeding

Select heterozygous individuals from F_2 population generated from cross between white and purple genotypes. Reject intense purple, if they have parent-type plant morphology. If the purple curding plants show difference in morphology, then use them for selfing to develop new lines. White curding plants should be rejected in

every generation. In purple broccoli, the recurrent selection was carried out in exotic segregating material and Delhi Purple Broccoli-1 line was selected, which was identified as Pusa Purple Broccoli-1 for cultivation during winter months in Delhi region. It is rich in health beneficial glucosinolates + anthocyanin (30.31 ± 0.68 mg/100 g fw). The Pusa Purple Broccoli-1 is a short-duration (75–85 days) variety suitable for taking early crop in purple broccoli. It will expand harvesting period of purple broccoli (December–January) since presently available variety Palam Vichitra is available during February–March months in plains. Unlike Palam Vichitra, DPB-1 has subtropical flowering habit, so seed production is possible in North Indian plains during winter season (IARI, 2020). The seeds produced in North Indian plains can be used for raising the crop and microgreens.

Backcross Breeding

It is useful to transfer *Pr* locus in a desirable background. For this, *Pr* locus itself acts as marker (due to variation in intensity of purple color of curd among homozygous and heterozygous varieties). Hence, backcrossing need to be carried out up to BC_2 or BC_3 generation followed by selfing/sibbing of horticulturally important individuals, having intermediate purple curding color trait, to derive homozygous lines for *Pr* locus. This may be done in later stages if the background/recurrent genotypes are from different group. It is a semidominant gene and its transfer is not like recessive or dominant gene. So, selfing of backcross population is required to isolate suitable lines.

Hybridization Method

In purple cauliflower, a new open-pollinated variety Pusa Purple Cauliflower-1 was developed through hybridization method using a cross between Pusa Snoball Kt-25 x Graffitti. It forms marketable curds during January–March months (IARI, 2020). A similar line of work is being undertaken by the authors for development of anthocyanin-rich purple cauliflower in early maturing Indian cauliflower.

Hybrid Selection

The *Pr* locus has incomplete dominance; hence, to attain intense colored curds, it is essential to have both parents in homozygous state for the *Pr* locus to get desired color in curds. The steps in breeding of purple color hybrids in cauliflower are shown in Fig. 19. Pusa Red Cabbage Hybrid-1 of cabbage was developed using red cabbage CMS lines and red cabbage fertile line as pollen parent (IARI, 2020). A CMS-based red cabbage hybrid has purple-green outer leaves and head is completely purple, very compact, and round. It matures in 70–75 days after transplanting. Average head weight is 1.2 kg and yield is 47.0 t/ha. The varieties developed in vegetable Brassicas for anthocyanin content are given in Table 1.

4.3.5 Breeding for Glucosinolates

Breeders of *Brassica* have sought to enhance the nutritional profiles of their plants by increasing the concentration of GSL, which includes glucoraphanin, glucoiberin, glucoerucin, and the indolic compound glucobrassicin. The development of broccoli

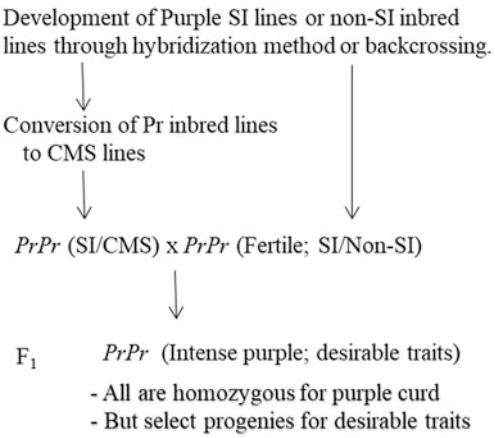


Fig. 19 Steps in development of purple color hybrid in cauliflower

with a higher glucoraphanin concentration as a result of extensive epidemiological data linking the therapeutic effects of broccoli to the bioactivity of sulforaphane, the ITC obtained from glucoraphanin, serves as an example of traditional breeding in vegetable *Brassica* crops. This development occurred as a result of extensive epidemiological data linking the nutritional benefits of broccoli to the biological processes of sulforaphane, which was produced from glucoraphanin (Riso et al. 2014). Sulforaphane may have anticarcinogenic characteristics, which are beneficial to human health, as shown by a variety of in vitro and animal studies (Mithen et al. 2000; Juge et al. 2007). In the 1990s, researchers from the UK evaluated nine distinct wild species of *B. oleracea* and discovered that *Brassica villosa* species had the greatest amounts of glucoraphanin out of all the species. According to Faulkner et al. (1998), hybrids produced by mating broccoli inbred lines with this wild cousin species had higher levels of glucoraphanin and were better able to stimulate phase 2 enzymes than broccoli inbred lines. This was found to be the case even though broccoli inbred lines had higher levels of glucoraphanin. Later on, Mithen (2003) made the discovery that the long-term introgression of two *B. villosa* L. chromosomal regions into broccoli led to an increase in the amount of ITC that was produced by the plant. One region of the B. villosa genome positioned on LG 5 increased 3C GSLs, while other regions placed on LG 9 and 2 elevated GSLs emanating from methionine. Because nitrile synthesis is inhibited in these genotypes, not only do they have higher GSL levels, but they also have a greater capacity for converting GSL into ITC. The *MYB28* allele, which was introgression from the *B. villosa* region, was shown to be highly expressed in hybrids with high levels of glucoraphanin. The amount of glucoraphanin found in these hybrids is about two to three times higher than that seen in conventional hybrids. One of the early uses of these results was the granting of a patent in the USA for broccoli sprouts that were created using genotypes that were rich in GSL (Fahey et al. 1997). Beneforté® broccoli is the brand name given to glucoraphanin-packed broccoli hybrids that are grown commercially in the Netherlands and the USA. A human intervention study

made use of the experimental hybrid HG1, which was a high-glucoraphanin F_1 hybrid originating from the UK initiative.

Some types of *Brassica* plants have benefited from additional breeding and selection methods carried out by GSL. For instance, a strategy of boosting marrow stem kale with a low indole GSL level that used a full-sib family selection approach proved successful (Bradshaw and Wilson 2012). In contrast, plant breeding use divergent mass selection to establish populations of individuals with genetically identical ancestry and outstanding GSL attributes. This is accomplished via the breeding of Brassicas. According to Stowe and Marquis (2011), this kind of choice on a speedily *B. rapa* variety can impact the cumulative GSL of the leaves. The Misión Biológica de Galicia (MBG-CSIC) selected for both high and low amounts of glucoiberin, sinigrin, and glucobrassicin, in the leaf tissue of turnips leaves (*Brassica rapa*). These selection programs were also used on the kale's ancestral population (Stowe and Marquis 2011).

5 Genetic Diversity Analysis

Brassica vegetables were originated from a common ancestor *B. oleacea* var. *sylvestris*. During evolution major and minor mutations as human selection for adaptive and economic traits resulted in diverse botanical varieties in *B. oleracea,* namely var. *caitata* (cabbage), var. *acephala* (kale), var. *botrytis* (cauliflower), var. *gemmifera* (brussels sprouts), and var. *italica* (broccoli). These botanical varieties have different plant parts as edible portion, hence differ in composition of quality constituents. *Brassica* species also have great extent of diversity in phytochemicals and metabolites (Branca et al. 2018). They found a wide range of carotenoids, total anthocyaninascorbic acid, glucosinolate concentration, and polyphenols, some of which might be employed as phytoconstituents indicators for accountability. Wild *Brassica* species contain higher levels of glucosinolates than the *B. rapa* subsp. *oleifera*, *B. oleracea* and *B. rapa* (Fig. 20). They also reported inter- and intraspecies variation in *Brassica* species for total and individual glucosinolates such as sinigrin, gluconapin, glocobrasicanapin, progoitrin, glucoiberin, glucoraphanin, and glucobrassicin. They reported only glucoiberin (100% of total GSLs) in *B. rupestris* and predominantly progoitrin (86% of total GSLs) in Pop. 1 (Sadinia) of *B. insularis*. In wild *B. oleracea*, gluconapin (79%) was the predominant GSL followed by glucobrassicin (18%) and progoitrin (3%), while in the case of cultivated *B. oleracea*, predominant GSLs in *B. o.* var. *capitata* were glucobrassicin (65%), sinigrin (25%), and I-methoxyghrcobrassicin (15%). Glucobrassicin (57%) was also highest GSL in *B. o.* var. *botrytis,* while in *B. o.* var. *italica*, glucoraphanin (48%) was major GSL followed by glucobrassicanapin (39%). The GSLs profiles of cauliflower does not matche with its probable ancestor *B. cretica*, which had gluconapin (86–97% of total GSLs) as the predominant GSL, indicating for major changes in the GSL pathway genes during evolution process. The study of GSLs in 51 species of Cruciferae from Middle Eastern region detected 30 compounds with three unknowns of limited occurrence. They found *p*-hydroxybenzylglucosinolate in six species, 4-methylsulfinyl-3-butenyl, 3-methoxycarbonylpropyl,

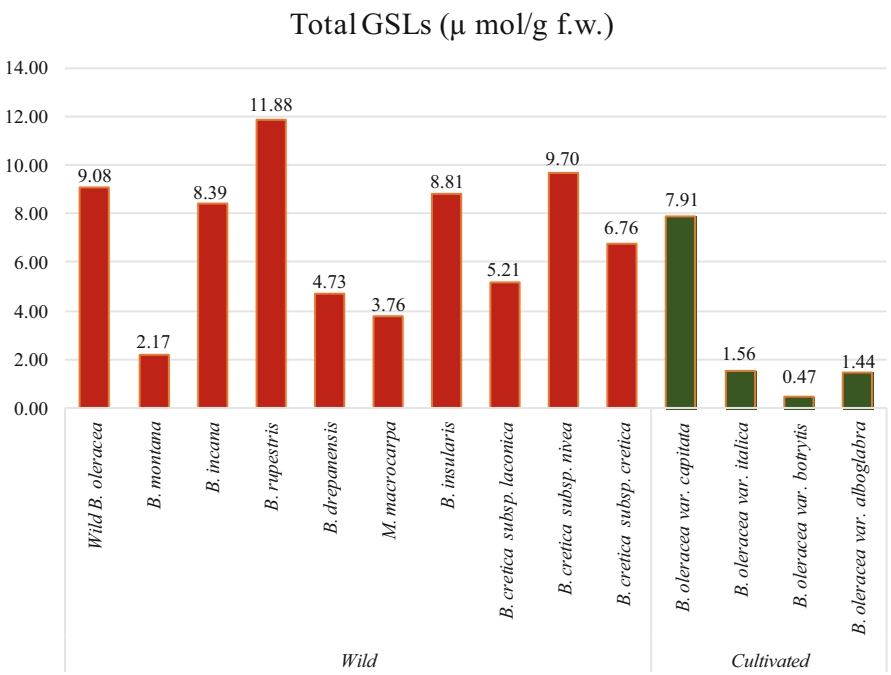


Fig. 20 Total glucosinolates content in cultivated and wild vegetable *Brassica*

and *p*-rhamnopyranosyloxybenzyl in *Sinapis aucheri, Erysimum*, and *Thlaspi perfoliatum*, respectively. They concluded that the distribution of GSLs is very useful at the species and genus levels but may provide only minimal support to the tribal alliance of certain genera. It supports the removal of *Sinapis aucheri* from the genus, the recognition of *Fibigia clypeata* and *F. macrocarpa* as distinct species, and the alliance of *Ochthodium* with *Euclidium* and *Arabis* with *Drabopsis.*

Glucosinolates component variability is associated to the A, B, and C genomes of three ancient vegetable *Brassica* species having diploid genetic chromosomes. These genomes are from species of vegetable *Brassica*.

- *B. nigra (BB, 2n = 16)* has glucosinolates with 3C side chains, which are the result of a single elongation process.
- The glucosinolates in *B. oleracea* (CC, 2n = 18) have either 3C or 4C side chains.
- The glucosinolates in *B. rapa* (AA, 2n = 20) have either 4C or 5C side chains.
- *Digenomic species B. carinata* (BBCC, 2n = 34), *B. juncea* (AABB, 2n = 36), and *B. napus* (AACC, 2n = 38) possess following glucosinolate composition that consists of the profiles of two elementary species.
  - Glucoiberin: white cabbage, red cabbage, cauliflower, and kale
  - Sinigrin: *B. oleracea* (cabbage, kale, brussels sprouts, cauliflower, and knol khol), *B. juncea*, and *B. carinata*
  - Glucoerucin: *Eruca sativa*

- Glucoraphanin: *B. oleracea* (broccoli, cauliflower, and kohlrabi) and *Eruca sativa*
- Gluconapin: *B. rapa* (bok choy and turnip) and *B. oleracea* (broccoli)
- Progoitrin: *B. rapa* (turnip and pak choi) and *B. oleracea* (brussels sprouts, cauliflower, and broccoli)
- Dehydroerucin: radish (80% of all glucosinolates)
- Glucobrassicanapin: *B. rapa* vegetables

Assefa et al. (2019) analyzed intact glucosinolate in the leaves of 50 germplasm collections and commercial cultivars of *B. rapa*, *B. juncea*, and *B. oleracea* from six countries grown under uniform cultural conditions. They reported total GSLs content in the range of 36.80–2383.12 μmol/kg DW with predominance of aliphatic GSLs (23.0–98.9%). Gluconapin and glucobrassicanapin contributed the greatest proportion.

6 Molecular Mapping of HR Genes and QTLs

6.1 Carotenoids

6.1.1 Marker-Assisted Breeding of *Or* Gene

The *Or* gene is a semidominant gene and it is almost impossible to differentiate between white and orange individuals as well as homozygous orange and heterozygous orange individuals because these all produce normal seedlings and show similar vegetative growth pattern. However, in later phases some visible morphological differences occur because of the impact of homozygous orange gene such as enhanced petiole length, small curds, and delayed flowering in case of homozygous orange (Crisp et al. 1975). To differentiate white (*oror*) and orange (*OrOr/Oror*) individuals, Li and Garvin (2003) identified ten amplified fragment length polymorphism (AFLP) markers using bulked segregant analysis. They converted four AFLPs most close to the target *Or* gene into restriction fragment length polymorphism (RFLPs). Mapping of these RFLPs could find close linkage between two markers and the target *Or* locus. Of these, one marker was at 0.5 cM from the *Or* locus while another RFLP was on another side of the locus at a distance of 1.6 cM. Subsequently, they converted two close flanking RFLPs into SCAR (sequence-characterized amplified region markers). It has been discovered that the SCAR markers (SA4, SA7, and SA9) are associated to the *Or* region. SA4 marker was found to be gene-specific marker for foreground selection. These markers are useful for large-scale screening of *Or* genotype in segregating population from the white individuals.

A single-locus recessive mutation called the *Or* in Chinese cabbage results in an accumulation of carotenoids in the plant's head leaves, giving them an orange coloration (Zhang et al. 2008). Microspore culture was used to develop a doubled haploid population from an F1 cross between the lines "91–112" and "T12–19" (white head leaves × orange head leaves), which revealed the SCAR markers connected to the *Or* gene. In Chinese cabbage, the Or gene is responsible for producing orange head leaves. Three RAPD markers, P67M54-172, OPAX18-656,

and OPB01-845, were effectively converted into the SCAR markers SCOR 127, SCOR204, and CSR 845. These three SCAR markers were used in a genetic analysis to map to the same linkage group as the *Or* gene with a LOD score of 6.0, implying that the *Or* gene should be found on A genome linkage group R9. They were also combined with two previously identified SSR markers, Ni4D09 and BRMS-51 (on linkage group R9). Furthermore, accuracy values of 92%, 90%, and 89.1% were found when 110 distinct inbred breeding lines of Chinese cabbage were examined using these three SCAR markers. This shows that these markers might be used for marker-assisted testing in the selection of beta-carotene-rich genotypes.

6.2 Anthocyanins

Ken et al. (2010) identified five RAPD markers around the locus for purple pigmentation in *Brassica rapa*. With the use of two SSR markers, one CAPS marker, and four RAPD markers, a preliminary linkage map of the pigmentation locus was constructed. A 0.84-kb segment of the RAPD marker OPU10 was amplified, cloned, sequenced, and then turned into a CAPS marker (OPU10C). On the partial linkage map and in linkage group R07 of the previously published reference map of *B. rapa*, this CAPS marker was mapped 4 cM distant from the pigmentation locus. Zhang et al. (2016) discovered a mustard (Anm) anthocyanin locus on *B. rapa* chromosome A02. The amplified bits were sequenced and aligned, and they revealed that purple Chinese cabbage contains purple mustard components on chromosome A02. Five genes were expressed at higher levels in purple Chinese cabbage than in nonpurple Chinese cabbage, according to the expression patterns of 12 anthocyanin-related genes on A02. Liu et al. (2021) found two QTLs linked with broccoli anthocyanin production on chromosomes 7 (*BoPur7.1*) and 9 (*BoPur9.1*). They then performed high-resolution mapping of *BoPur9* and discovered that the key anthocyanin biosynthesis QTL region is 73 kb in size and comprises 14 genes. AT5G07990 (gene encoding flavonoid 3′ hydroxylase), a putative gene for *BoPur9.1*, had a considerable degree of similarity with *Bo9g174880, Bo9g174890*, and *Bo9g174900.* The expression of BoF3'H was much higher in BT 126 than in SN60. Chiu et al. (2010) mapped the *Pr* locus in purple cauliflower at 0.3 cM resolution, including *BoMYB2, BoMYB3, BoMYB4,* and the *Pr-D* locus. Among these markers, *BoMYB2* has been shown to be located inside the *Pr-D* locus. The expression of anthocyanin regulatory and biosynthetic genes was shown in eight marketed cultivars (Yuan et al. 2009). They discovered that all kinds of red cabbage have a similar mechanism for controlling anthocyanin production. Except for CHI, all other structural genes, including *CHS, F3H, F3'H, DFR, LDOX,* and *GST*, were constitutively upregulated in red types, which coincided with an increase in the transcript levels of a *bHLH gene, BoTT8,* and a *MYB* transcription factor, *BoMYB2*.

6.3 Glucosinolates

The process of glucosinolate synthesis in *Arabidopsis* and *Brassica* is well established, as previously said, particularly as a consequence of investigations

carried out over the past ten years. It has been clear since the 1970s that glucosinolates are important for both human and animal nutrition. In the 1990s, preliminary efforts were made to identify the region of the genome that was responsible for GSL formation. One of these approaches included searching at quantitative trait loci (QTLs). In the last few decades, several researches have tried to identify QTLs related to GSL inheritance due to the accessibility of molecular markers.

Breeders may choose certain genotypes by predicting the parental alleles at the QTLs from the genotype at marker loci connected to those QTLs (Uzunova et al. 1995). Finding QTLs is also the first step in finding and cloning the genes responsible for the desired traits. The accessibility of the genetic data of the major *Brassica* crops, similar to *Brassica* species, and understanding of the genes that govern GSL formation in *Arabidopsis* have facilitated the search for possible loci for GSL synthesis in *Brassica* crops. QTL analysis has been regularly used in *Brassica* crop breeding to increase GSL content. Numerous studies on QTLs have been conducted on the vegetable forms of *B. rapa*, such as Chinese cabbage, pak choi, turnips, turnip tops, and cima di rapa. Lou et al. (2008) identified QTLs for GSL amounts and deposition in *B. rapa* young leaves in two inbred populations comprised of Kairiyou Hakata, pak choi, a Japanese turnip, and yellow sarson cultivars. Aliphatic GSL concentration was regulated by 16 QTLs, indolic GSL concentration by 3, and aromatic GSL concentration by 3 QTLs. To find candidate genes, metabolomics have often been utilized in conjunction with genomic, transcriptomic, and other platforms. Del Carpio et al. (2014) examined the profile of GSLs and the transcript quantity of genes associated with GSL synthesis in the leaves of plants from a breeding population generated from a combination of a pak choi and yellow sarson plant. In genomic areas on linkage groups A03 and A09, significant microsatellite quantitative trait loci (mQTLs) were generally co-localized. The results of the 94 probes were analyzed, and 42 of them indicated at least one eQTL. These findings depict 25 candidate genes. Both cis- and trans-eQTL were present in the majority of hotspots in the genomic areas at A09 and A03 for these genes, with the majority of genes exhibiting trans-eQTL effects at the locations of *MYB28* in A09 and *MYB29* and *UGT74B1* in A03. Nuclear magnetic resonance (NMR) was used by Bagheri et al. (2013) to investigate the genetic variation for a variety of secondary metabolites in *B. rapa*. Six GSL QTLs were assigned to the loci A10, A9, A5, and A3, with two QTLs in A3 and two QTLs in A5 presumably co-locating. The progoitrin QTL and the MYB29/MYB28 map sites were detected on A3 at 95 cM. Recently, efforts have been made to cultivate *B. oleracea* plants with high glucoraphanin contents. In a broccoli mapping population, Brown et al. (2015) found 14 QTLs for both general and particular GSLs. While *GSL-PRO* and *GSL-ELONG* have been linked to specific loci in *C05* and *GSL03*, respectively, *GSL-ALK* has been proposed as the gene responsible for GSL12 variation in C09. Sotelo et al. (2014) identified 18 metabolic QTLs for total and individual GSL content in leaves, flower buds, and seeds, in a inbred population generated from an intercross of broccoli and Chinese kale inbred lines. It was believed that the GSL-OH, GSL-PRO, and GSL-ALK genes were responsible for the QTL-3, QTL-5.1, and QTL-9.2, modifications, respectively. The indolic GSL pathway genes ATR1, CYP81F2,

CYP79B3, and CYP79B2 were among the candidates for QTLs QTL-8.1, QTL-2.1, QTL-7.4, and QTL-1.2, respectively.

6.4 QTL Analysis for Quality Traits in *Brassica* Vegetables

There has been an increase in consumer demand for nutrient- and biochemically dense foods that have a positive effect on human health during the past decade. Brassicaceae plants include crucial natural physiologically active compounds, such as enzymes, pigments, vitamins, and particular secondary metabolites, with antioxidant, cardioprotective, anticarcinogenic, and anti-inflammatory properties and immune-boosting properties (El-Mogy et al. 2019; Artemyeva et al. 2014). Biparental quantitative trait locus (QTL) linkage mapping is an efficient method for finding genetic regions that co-segregate in the trait of interest within the research population. Mapping quantitative trait loci (QTL) has shown to be a useful technique for analyzing complex traits controlled by several genes (Salvi and Tuberosa 2005). Thousands of quantitative trait loci (QTL) that control multiple health-promoting characteristics have been identified in varied species by linkage mapping techniques (map-based QTL), and dozens of important genes have been cloned by fine mapping of QTL in several crop species (Bernardo 2008). In populations of doubled haploid lines of *Brassica rapa*, Egorova et al. (2021) discovered a correlation between molecular markers and chromosomes and the amount of dry matter and physiologically active compounds such as sugars, carotenoids, anthocyanins, ascorbic acid, and chlorophylls a and b. They identified 102 quantitative trait loci (QTLs) that regulate the expression of biochemical quality parameters under both short- and long-daylight circumstances. In addition, they discovered that the majority of loci controlling all of the investigated biochemical parameters were located in the fifth, sixth, seventh, and ninth linkage groups, which were consistent with field and controlled environmental data. The outcomes of prior analyses of the most crucial biochemical quality criteria in Brassicaceae have been presented. Artemyeva et al. (2014) identified quantitative trait loci (QTLs) for five nutritional quality markers, namely beta-carotene, ascorbic acid, total protein, and chlorophylls a and b, using two mapping populations of *Brassica rapa* doubled haploid lines. QTLs, their effects, the fraction of phenotypic variability impacted by each QTL, and molecular markers genetically related with each QTL were discovered for each characteristic investigated. Significantly, Jan et al. (2016) identified a significant variation in the cumulative seed protein content of several *Brassica rapa* ecotypes, which has significance for agricultural development and effective utilization. Ullah et al. (2017) examined the heritability, genetic variation, and association of several biochemical characteristics in advanced populations of *Brassica rapa*. They discovered substantial differences in glucosinolate, oil content, protein content, oleic acid, linolenic acid, and other biochemical indicators. Zhao et al. (2007) mapped the *B. rapa* QTL for phytate and phosphate levels in seeds and leaves. Lou et al. (2008) found the QTL for the glucosinolate aggregation trait, which influences the nutritional value of *B. rapa* and is a vital plant defense mechanism. The location of QTLs

linked with the amount of sugars, amino acids, organic acids, aromatic compounds, and glucosinolates in seedlings was found using QTL analysis of metabolites (Bagheri et al. 2013).

Leaf heads of cabbage (*Brassica oleracea*) and Chinese cabbage (*B. rapa*) are vital sources of minerals, crude fiber, and vitamins in the human diet. Yu et al. (2013b) conducted QTL mapping on 150 RILs derived from the combinations of heading and nonheading Chinese cabbage in order to investigate the genetic component of leafy head in Chinese cabbage (*B. rapa*). The parental genome resequencing identified more than one million SNPs. Genotyping RILs with high-quality SNPs provided a high-quality genetic map with 2209 markers and 18 QTLs (LOD of more than three and phenotypic effect [R2] of more than 5% for six head features, from which three candidate genes were identified). Gao et al. (2022) analyzed 16 important agronomic characteristics in 240 F_2 individuals derived from two separate inbred lines of Chinese cabbage, ZHB and G291. A genetic map of 105 intragenic simple sequence repeat (ISSR) markers dispersed over ten linkage groups (LGs) spanning over 2034.1 cM in length was constructed having an average interlocus distance of 21.75 cM. His team discovered 48 QTLs on the studied LGs for the assessed major agronomic parameters, with LOD values varying from 2.51–12.49, which accounted from 3.41 and 26.66% of the morphological characters.Several quantitative trait loci (QTLs) for anthocyanin accumulation in Chinese cabbage have been found (Guo et al. 2015). In an F_3 population, genes regulating basic hereditary traits, such as seed color, seed erucic acid content, and the presence of leaf hairs, have also been identified (Teutonico and Osborn 1994). Bagheri et al. (2013) revealed QTLs influencing seedling metabolite and seed tocopherol concentrations in a RILs generation of *Brassica rapa* obtained from L58 (with a higher level of glucosinolates and phenylpropanoids) and R-o-18 (having high sucrose, glucose, and glutamate levels). The QTLs that were linked with seed tocopherol concentrations (a-, ß-, d-, α/β, and total tocopherol) were identified on chromosomes A10, A9, A6, and A3, and they accounted for 11–35% of the variance that was significant. The region on A3 is located near to the *BrVTE1* gene, which codes for tocopherol cyclase. NMR spectroscopy reveals that seedlings contain organic/amino acid, sugar/glucosinolate, and aromatic compounds. The majority of identified chemicals have their QTL locations established. In contrast to prior investigations, specific loci for glucosinolate amounts have been identified. Five to seven QTLs with additive and/or epistatic effects have been discovered in a doubled haploid (DH) population of *B. napus* for tocopherol, total tocopherol, and their ratio (Marwede et al. 2005). In another DH population of *B. napus* and its rebuilt F2 progeny, 50 QTLs and related markers for tocopherol composition and content were identified (Wang et al. 2012b). In two recombinant inbred line (RIL) populations of the *Brassica* reference plant species *Arabidopsis thaliana*, 14 QTLs affecting seed tocopherol content and composition have been found. Using mutational analysis and genomic-based methodologies, several researches have discovered and cloned the genes of the tocopherol synthesis pathways in *A. thaliana* (Bagheri et al. 2013).

Numerous secondary plant metabolites in the plant metabolome may have a nutritional effect. Glucosinolates are sulfur-containing plant compounds that have anticancerous properties. They encompass more than one hundred plant secondary

metabolites, the majority of which are found in the Brassicaceae family. Glucosinolates are of particular interest because of their anticarcinogenic properties (Hayes et al. 2008). In two doubled haploid (DH) populations of *B. rapa*, leaf aliphatic glucosinolate QTL mapping identified 16 loci regulating aliphatic glucosinolate concentration (Lou et al. 2008). According to comparative genomic studies, 102 putative glucosinolate-producing genes in *B. rapa* are orthologs of 52 associated genes in *A. thaliana* (Wang et al. 2011). Five consensus areas align to chromosomes C1, A9, A2, C3, and A8 in the genome of *Brassica napus* Darmor-bzh. In addition, 24 orthologs of functional key genes for biosynthesis of fatty acids were detected.

7 Marker-Assisted Breeding for HR Traits in *Brassica* Vegetables

7.1 Carotenoids

Kalia and his colleagues showed a marker-assisted breeding technique for the efficient introgression of the *Or* gene in Indian cauliflower for the purpose of beta-carotene biofortification (Fig. 21). Repetitive backcrossing and screening of advanced items employing accessible foreground markers were utilized in this approach. They employed background markers to quickly retrieve the genome of

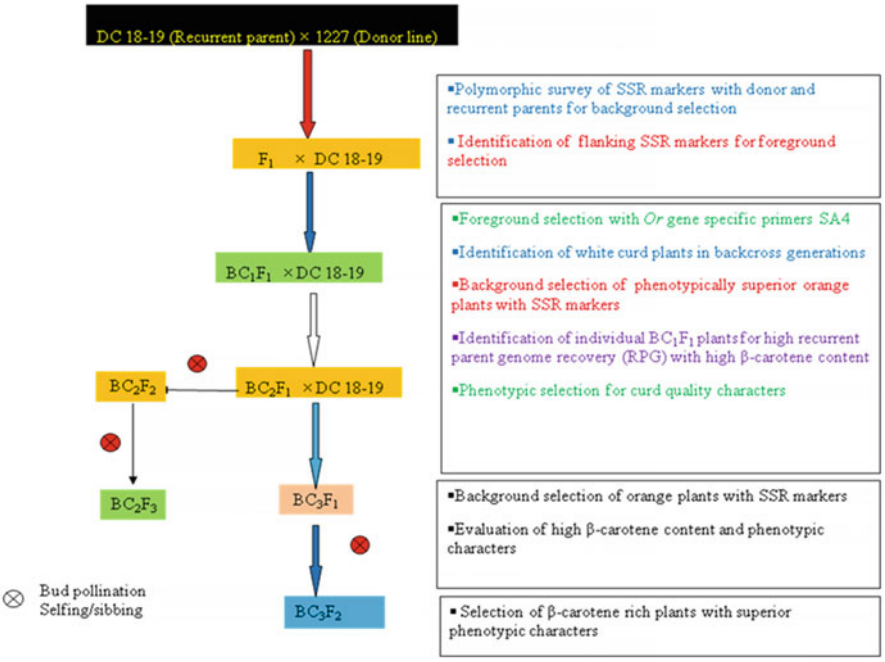


Fig. 21 Marker-assisted breeding of *Or* gene in Indian cauliflower. (Source: Muthukumar 2016)

the 'DC-18-19' recipient genotype. Kalia et al. (2020) used PCR to confirm the AFLP-based SCAR marker SA4 in *Or*-introgressed lines of Indian cauliflower. They also reported the creation of a functional SSR marker that allows the selection of *Or* gene carrier plants at the seedling stage. Furthermore, 400 SSR markers encompassing nine chromosomes were employed for polymorphism survey and background selection. On this basis, 35 polymorphic markers were discovered and employed for background selection in selected BC2 plants in order to extract the parent genome.

7.2 Anthocyanin

In purple cauliflower, Chiu et al. (2010) found a high-resolution genetic mapping of the *Pr* was done having *BoMYB2, BoMYB3, BoMYB4,* and the *Pr-D* locus at 0.3 cM. Of these markers, BoMYB2 is shown to be located within the *Pr-D* locus.

8 Map-Based Cloning of HR Genes/QTLs in *Brassica* Vegetables

Li and Garvin (2003) sought to isolate *Or* using a map-based cloning technique, utilizing the cauliflower *Or* mutant as a model, in order to elucidate the molecular mechanism through which the *Or* gene controls carotenoid accumulation. They used a flanking marker technique in which they used two previously identified, locus-specific, SCAR markers for the study of 1632 F_2 individuals to create a high-resolution genetic linkage map of the *Or* locus area (Li et al. 2003). They defined the *Or* locus to a 50-kb BAC clone, which corresponds to a genetic interval of 0.3 cM, utilizing positional cloning using BAC clones.

GSL-ALK was cloned and tested in two recombinant inbred line populations of *B. oleracea*. Table 2 depicts the evolutionary variety in glucosinolates. Giamoustaris and Mithen (1996) proposed a model in which 3-methylthiopropyl GSL is sequentially transitioned to 3-methylsulfinylpropyl and finally to 2-propenyl GSL by the action of dominant alleles at two distinct loci in a backcross inbred population (BILs) generated from a cross between two *B. oleracea* relatives viz. GSL-ALK and GSL-OXID. These two loci were assigned to the same LG as the *B. napus* LG 19 loci, which were validated using RFLP mapping in *Arabidopsis* but also wild *Brassica* species. On the basis of these results and previous research on *Arabidopsis*, four key biosynthetic loci were proposed: GSL-AOP, ELONG, OH, and OX. All four of these loci are prefixed. *Brassica* includes glucosinolates with three, four, and five carbons, but *Arabidopsis* includes glucosinolates that have as many as eight carbons. Secondary metabolites identified in broccoli include carotenoids, glucosinolates/sulforaphane, flavonoids, and phenolic acids. Multiple loci/genes have been found that influence the accumulation of these compounds in broccoli. Broccoli may be investigated using *Arabidopsis* genetic models of secondary metabolite synthesis (Frerigmann and Gigolashvili 2014). Some broccoli genes,

including *cytochrome P450 83A1 (CYP83A1), cytochrome P450 79F1 (CYP79F1), sulfotransferase UDP-glucosyltransferase 74B1 (UGT74B1), 18 (ST5b),* and *flavin-containing monooxygenase GS-OX1 (FMOGS),* are isolated directly using homologous cloning (Rahim et al. 2019). Furthermore, genetic mapping was carried out in order to discover genetic loci that impact the variability of secondary metabolites. Sotelo et al. (2014) used genotyping to identify 82 considerable QTLs for specific and quantitative glucosinolate production in leaves, seeds, and flower buds, with QTL9.2 (proposed candidate as GSL-ALK) being one of the important in glucosinolate differences and implying epistatic interrelations with other loci (Sotelo et al. 2014). Using a broccoli mapping population, Brown et al. (2015) developed a genetic linkage map, discovered 14 QTLs associated with indolic, aliphatic, or aromatic glucosinolate storage in florets, and showed that a locus GSL12 on C09 is responsible for almost 40% of progoitrin phenotypic diversity. Li et al. (2021b) used a DH population to do genetic mapping for sulforaphane metabolism in broccoli florets; they revealed 18 QTLs for sulforaphane synthesis in broccoli flower heads, six of which were found in several environments. Brown et al. (2015) utilized a known function population to construct a functional genomic map and an SNP array to detect three quantitative trait loci (QTLs) involved in carotenoid variation in broccoli heads (Brown et al. 2014). In at least two studies, Gardner et al. (2016) utilized a QTL analysis saturated with SNP markers on an Illumina 60 K array to identify 23 loci for overall phenolic content and its constituents in the population, which Brown et al. (2014) studied in high-density single nucleotide polymorphism (SNP) array mapping in Brassica oleracea for identification of QTL associated with carotenoid variation in broccoli florets. In the BolTBDH mapping population, a total of 33 QTLs that influence the amount of phenolic compounds found in seeds, leaves, and floral buds were discovered (Francisco et al. 2016; Gao et al. 2014; Lee et al. 2017). Additionally, broccoli seeds, sprouts, and by-products were subjected to transcriptome analyses to discover genes with expression levels associated to glucosinolate synthesis.

9 Genomics-Assisted Breeding for HR Traits in Vegetable Brassicas

Using association mapping, linkage disequilibrium (LD) mapping, and genome-wide association study (GWAS) is useful for elucidating the molecular genetic foundation underpinning natural phenotypic variation. Through this several causal allele(s)/loci that had not been found in QTL mapping populations were identified. This potential technique represents a significant advancement in genetic research and has clearly shown to be a beneficial tool in the discovery of candidate genes. Association mapping is increasingly being employed in wild populations of Brassicaceae with high genetic variation to identify DNA-based markers linked to essential health-promoting features.

Using 60 K SNP array in the *Brassica*, Qu et al. (2017) performed a genome-wide association investigation of fatty acid content in 520 different genetic oilseed

germplasms. Using the PCA + K model in TASSEL 5.2.1, the researchers analyzed 62 genetic loci that were associated significantly with the substance of seven fatty acids, in as well as five consensus regions that pertained to the C1, C3, A9, A8, and A2, chromosomes of the *Brassica napus* Darmor-bzh genome, respectively. The researchers then discovered various orthologs of functional candidate genes involved in lipogenesis, excluding BnaA.FAE1 and BnaC.FAE1 on the A8 and C3 homologous genome blocks, which are recognized to have significant functions in the fatty acid biosynthetic pathways, as well as promising orthologs of these genes (*KCS17, CER4, LPAT4, FAB1, LACS9, KCR1, TT16,* and *ACBP5*).

One of the most essential features of oil crops is seed oil content, which acts quantitatively (Zhao et al. 2005; Delourme et al. 2006). Linkage mapping of several populations revealed that many QTLs regulate oil content in rapeseed (*Brassica napus*) (Qiu et al. 2006). However, the number of parents utilized in prior genetic linkage mapping of QTL represents a relatively tiny part of rapeseed germplasm, and it is unknown how often the QTL may be discovered again in practical breeding (Mackay et al. 2009). By using association analysis or association mapping, QTL may be found in a way that is independent of linkage mapping populations. Zou et al. (2010a) investigated the efficiency of association mapping for seed oil content in new-type *B. napus* lines and a collection of traditional B. napus varieties from various countries. The new-type *B. napus* has relatively rich genetic diversity, although it has a greater amount of linkage disequilibrium than the classic *B. napus*. Similarly, in the population of new-type *B. napus*, there was a higher diversity in oil content and a greater number of related markers. Meanwhile, more than half of the genetic loci to which the linked markers belonged were found earlier in linkage mapping research, demonstrating the effectiveness of association mapping in *B. napus*.

10 Recent Concepts and Strategies Evolved in *Brassica* Vegetables

10.1 Genome Editing

The technology of gene editing provides essential technological foundations for plant functional gene research and agricultural genetic modification. In recent years, CRISPR/Cas9-mediated gene editing has emerged as a promising experimental technique for crop genomic sequence editing, and a number of different crops have widely used it (Tao et al. 2021). Murovec et al. (2018) developed a method for genome editing in a variety of *Brassica* species, such as *Brassica napus*, *Brassica oleracea*, and *Brassica rapa*. They identified the vernalization-determinant phytoene desaturase gene (PDS) and FRIGIDA (FRI) gene, which are both engaged in the carotenoid production pathway and are located at separate locations. Erucic acid (EA), which is known to have adverse health consequences, is one of the fatty acids that affect the edibility and processing qualities of vegetable oils. Low erucic acid (LEA) has been a *B. napus* selection characteristic for years. The enzyme fatty acid elongase 1 (FAE1) is required for the synthesis of EA. *B. napus* possesses two

functionally identical copies of FAE1 on chromosomes A08 and C03. Liu et al. (2022) used CRISPR/Cas9 to produce targeted mutations on these two homologous copies of BnaFAE1 in three *B. napus* accessions with high EA (>30%) and oil (>50%) levels. The EA content of the mutant BnaC03 was reduced by more than 10%. FAE1 (c03), while the combined mutation of *BnaA08.FAE1* and *BnaC03.FAE1* (a08c03), resulted in almost no EA and a variable rise in oleic acid content in three BnaFAE1-edited accessions.

CRISPR-Cas9 genome editing of *FAD2* has created rapeseed seeds with low linoleic acid and high oleic acid (Huang et al. 2020), providing a unique strategy for generating high oleic acid content oil crops. In addition, Ozseyhan et al. discovered that utilizing CRISPR technology to delete FAE1 might significantly reduce VLCFAs (Okuzaki et al. 2018).

Zhai et al. (2020) employed CRISPR/Cas9 and guide RNA to activate basic helix-loop-helix transcription factors (bHLH-TF) in a yellow-seeded *B. napus*. *BnaC09.TT8b* and *BnaA09.TT8* are transparent tests. The double mutant's enhanced seed oil level, greater protein content, and changed fatty acid (FA) composition were likely due to its thinner seed shell. Naturally occurring LOF mutations in these genes produce a phenotype characterized by yellow seeds in *B. juncea* and *B. rapa* (Li et al. 2012a, 2012b). The principal technique by which crops store phosphorus is phytic acid, which is antinutritional as it chelates minerals and limits their absorption. This is because phytases were unable to hydrolyze it into myo-inositol and free inorganic phosphate (Padmaja et al. 2014). Sashidhar et al. (2020) significantly enhanced the quality of canola meal by inactivating 15 paralogs of the canola ITPK gene family using the CRISPRCas9 technology. Karunarathna et al. (2020) used *Agrobacterium tumefaciens* with CRISPR/Cas9 reagents to improve seed oil content by changing of hypocotyl segments of WOSR cv. RS306 (and hence oil per acre) and chemical mutagenesis. Seed oil content was raised without impacting seed germination or vigor by introducing LOF mutations to the *Bnasfar5* (10%) and *Bnasfar4* (9.7–14.5%) genes late in the seed maturation phase.

Canola oil is mostly made up of free fatty acids, such as monounsaturated oleic acid and a 2:1 ratio of polyunsaturated linoleic and linolenic acid (Hu et al. 2006). Because of their improved shelf life and thermal stability, vegetable oils with a greater oleic acid content (C18:1) are appealing. In plants, the FAT ACID DESATURATION 2 (FAD2) gene encodes stearoyl-acyl carrier protein desaturase, which converts stearic acid (C18:0) to oleic acid (C18:1). Okuzaki et al. (2018) designed a *CRISPR/Cas9* vector that preferentially targeted BnaFAD2a in both the A and C genomes. While the amount of C18 polyunsaturated fatty acids decreased drastically, the amount of oleic acid generated by homozygous mutant plants increased substantially.

Huang et al. (2020) recently attempted to change all three active BnaFAD2 gene targets and found that although the oleic acid content of homozygous BnaA05.fad2a mutant crops rose considerably (6–16%), the C18 polyunsaturated fatty acid content declined dramatically. He et al. (2022) intended to modify a glucosinolate transporter gene with no genetic variations in order to create a cultivar with reduced seed glucosinolate for quality and resistance breeding of polyploid *Brassica napus* canola, the world's second-largest source of edible protein and oil meal.

10.2 Nanotechnology

Nanotechnology has proven several benefits in agricultural sciences (Prasad et al. 2014). Nanoparticles increase plant yield by interacting directly or indirectly with plants through the soil. They promote plant development by controlling nutrient transport or regulating the amount of micronutrients, and they give phytopathogen resistance (Kriti et al. 2020). It has been demonstrated that metal nanoparticles such as Ag, Au, TiO_2, ZnO, and iron enhance plant growth. As zinc is an important element involved in several metabolic processes, zinc oxide nanoparticles (ZnO NPs) have been utilized extensively (Hafeez et al. 2013). Numerous experts have researched the effect of ZnO NPs on a variety of agricultural plants, and their findings indicate that ZnO NPs promote plant development. There have been very few papers discussing the application of nanotechnology to boost health-promoting compounds in Brassicas. Mazumder et al. (2020) employed an alpha amylase enzyme to create and characterize zinc oxide nanoparticles (ZnO NPs), and then evaluated the nanoparticles' positive influence on the growth and development of *Brassica juncea*. According to *in silico* research, the amino acids lysine, glutamine, and tyrosine contained in the alpha amylase enzyme play a crucial role in the transformation of zinc acetate dihydrate to ZnO nanoparticles. *Brassica juncea*'s biochemical parameters and oxidative enzymes were compared with those of ZnO NPs-treated plants. According to the findings, nanoparticles might serve as an alternative to conventional, dangerous chemical fertilizers. Sohail et al. (2019) analyzed the effect of green synthetic NPs on two distinct cultivars of *Brassica napus*. When B. napus was treated to ZnNPs at concentrations of 5, 15, and 25 mg l-1, root and shoot length considerably increased. Utilizing nanoparticles dramatically enhanced plant germination and the production of secondary metabolites and antioxidant enzymes. ZnNPs enhanced the levels of chlorophyll, superoxide dismutase, and total flavonoid content (TFC), as well as antioxidant enzymes, while lowering the levels of total phenolic content.

Nanoparticle-mediated gene delivery has lately gained popularity; it appears to be superior to conventional biomolecular techniques since it improves the efficiency of transformation in several plant species for both transient and stable genetic changes. Nanotechnology technological advances have enabled us to transcend the limitations of conventional biomolecule delivery methods, and nanoparticles (NPs) appear potential for the species-independent passive transport of proteins, RNA, and DNA (Cunningham et al. 2018).

The amalgamation of nanotechnology into the process of generating genetically modified organisms (GMOs) is a powerful method that includes the employment of nanoparticles (NPs), the formation of an involved nanioparticle in the deal with biomodifier molecules (the CRISPR/Cas system), and the transportation of biomodifiers into plant cells (Abd-Elsalam 2020; Demirer et al. 2021). However, no reports have been published on nanoparticle-mediated biomolecule delivery for the enhancement of health-promoting substances in Brassicas, despite the fact that the incorporation of nanotechnology into existing molecular technologies could provide a forum for overcoming obstacles to the production of genetically modified plants and biotransformation in this field (Cunningham et al. 2018; Gad et al. 2020).

11 Genetic Engineering for HR Traits in *Brassica* Vegetables

The *Or* gene is a unique carotenoid gene mutation that generates a high concentration of β-carotene in orange cauliflower's edible curd. The *Or* gene promotes the accumulation of carotenoids in weakly pigmented or nonpigmented tissues, such as curd and pith. The *Or* gene has a negligible influence on leaf carotenoids, floral petal pigmentation, and chromoplast appearance. The *Or* gene transformants are capable of causing carotenoids accumulation by inducing the formation of large carotenoids sequestering sheets as a carotenoids deposition sink.

In recent years, biosynthesis/regulation-related genes have been altered to increase broccoli's anticancer glucosinolate/sulforaphane concentration (Kim et al. 2020). Upregulation of *BoMYB29* in the AG1012 (DH line) triggered the aliphatic glucosinolate pathway and increased the synthesis of glucoraphanin pathway, according to Zuluaga et al. (2019). Despite the absence of transgenic broccoli plants, this study provides a benchmark for increasing glucosinolate levels in broccoli (Li et al. 2019b). Despite research suggesting the significance of *BroMYB28* in glucoraphanin synthesis (Cao et al. 2021), its role in broccoli was not verified until 2019 (Kim et al. 2020). Broccoli accumulates glucoraphanin due to the transitory overexpression of *BroMYB28* by *Agrobacterium*. Recombinant broccoli was engineered by upregulating myrosinase, MAM1, and FMOGS-OX2, either alone or in combination. Genetic recombination increased sulforaphane content of individual myrosinase, MAM1, and FMOGS-OX2 by 3.7, 1.7–3.4, and 1.6–2.7 folds above wild-type plants, respectively, while the triple transgene increased it by 1.86–5.5-folds.

Numerous *Arabidopsis* and *Brassica* genes with comparable roles have been cloned and characterized, and the GSL synthesis pathway is well established. Genetic engineering is now being used to change the GSL content of cruciferous crops by concentrating on various phases of GSL production and degradation. Transgenic techniques that allow the overexpression or silencing of a single gene or a large quantity of genes enable for the rapid and precise manipulation of metabolic activities to obtain certain plant traits. The insertion of transgenes by genetic modification is a novel method for modifying the amount and type of GSLs. Liu et al. (2012) created glucoraphanin-rich B. napus seeds by inhibiting the *GSL-ALK* gene via RNA interference. Effective transfer of *Arabidopsis* genes is responsible for the generation of aliphatic and indolic GSLs into Chinese cabbage lines (Zang et al. 2008a, b). Overexpression of these genes led to the development of transgenic strains that accumulate indolyl (glucobrassicin and 4-methoxyglucobrassicin) and aliphatic (gluconapin and glucobrassicanapin) GSLs. Various strategies targeting biosynthetic or regulatory genes in the GSL biosynthesis pathway may be utilized to engineer GSL metabolism in plants. It appears that transcription factor regulation is more successful than genes encoding particular enzymes in governing metabolic pathways in plants. In *Arabidopsis*, around 20 genes with possible regulatory involvement in GSL metabolism have been discovered in recent years. In the near future, more research will provide the knowledge essential to simplify the complicated GSL biosynthesis in plants. Moreover, genetic modification has permitted the production of GSLs in

non-*Brassica* plants, such as *Nicotiana benthamiana* (Geu Flores et al. 2009), paving the path for GSL-enriched diets in other vegetable products, such as cancer-preventative commodities. Due to the success of breeding programs in reducing GSL levels in *B. napus* by conventional breeding and the rejection of genetically manipulated (GM) crops in certain oilseed rape–growing zones, the commercial sector is less inclined to explore GSL reduction through gene transformation. Likewise, consumers may dislike genetically modified broccoli with a high GSL content. Modern, guided genetic engineering is speedier, simpler, and more predictable than conventional breeding. The creation of a molecular key based on the genomic sequences of *Brassica* plants and *Arabidopsis* would speed up the marker-assisted discovery of target GSLs, therefore boosting the number of desirable GSLs in edible Brassicas.

12 Role of Bioinformatics as a Tool

Similar to other biological disciplines, nutrition science may gain a great deal from bioinformatics resources. If nutrition researchers, particularly those in the area of nutrigenomics, add bioinformatics abilities to their toolkits, they stand to gain a great deal from the many new opportunities that have emerged since the decoding of the human genome. Numerous bioinformatics techniques have the potential to significantly advance food science. For instance, bioinformatics sequence similarity algorithms have been used to find sweet and sour taste receptors in animals. Similarly, bioinformatics technology might be utilized to link diverse flavor components for the development of new goods based on customer preferences (Liu et al. 2008). There are several molecular marker techniques for detecting microbes that cause deterioration and illness. In addition, the Food and Drug Administration has developed a microarray-based tool for the molecular characterization of food-borne illnesses (Fang et al. 2010). In addition, bioinformatics tools may be utilized not only to determine the allergenicity and cross-reactivity of dietary proteins (Jenkins et al. 2005), but also to identify bioactive peptides (Walther and Sieber 2011). AllerMatch (Fiers et al. 2004), the Informal FARRP Allergen database, and SDAP (Mari et al. 2006) provide data on food allergens, while the protein sequences from UniProtKB, SwissProt, and TreMBL may be used to identify bioactive peptides. In addition to these bioinformatics tools, there are other databases with specialized neutrigenomics applications, such as Food DB, EuroFIR-BASIS, and the Foodomics database. FooDB (http://foodb.ca) is a comprehensive database on the chemistry and biology of food components. The EuroFIR-BASIS database, which is available at https://www.eurofir.org, combines biological activity and dietary composition for pharmacological activities in plant-based diets. FoodWiki database is a repository for consensus-based food and nutritional information, whereas Foodomics database is a repository for food molecular profiles (Kumar and Chordia 2017). Currently, it is obvious that advancements in "omics" approaches have not only led to the accumulation of massive datasets, but have also stimulated the creation of bioinformatics tools for analyzing these datasets. It is imperative that researchers in the nutrition

sciences utilize these analytical tools to mine information relevant to genomics, epigenomics, transcriptomics, proteomics, and metabolomics, as well as to investigate and decipher the composition of food, its microconstituents, nutritive value, and the chemistry and biology of food components.

12.1 Gene and Genome Databases

The genetics of *B. oleracea* have been researched throughout the last ten years. The entire genome of cabbage line 02-12 was made public by Liu et al. (2014). The whole genome of the cabbage double haploid mutant TO1000 was published by Parkin et al. (2014). Next technologies of genomic sequences were used in the construction of these two genomes. Third-generation sequencing technology was used in subsequent studies to finish the genomic construction and produce high-quality, rich genomes of variably shaped cabbage lines (D134, OX-heart, and JZS) (Belser et al. 2018), broccoli (HDEM) (Lv et al. 2020; Cai et al. 2020), and cauliflower (D134, OX-heart, and JZS) (Sun et al. 2019; Guo et al. 2021). The *Brassica* database was created by Cheng et al. (2011), and it is an online repository for extensive genomic and genetic data on important *Brassica* plants. *Brassica rapa* was the first genome to be sequenced in its entirety, and those results were used to create BRAD (Chiifu-401-42). It contains the complete *B. rapa* genome sequence, which was constructed from BAC clone sequences, and Illumina GA II short reads as well as predicted genes and related annotations, transposable elements (TE), noncoding RNAs, *B. rapa* genes orthologous to those in *A. thaliana*, linkage maps, and genetic markers. Unlike BLAST and Gbrowse, BRAD's search and data mining capabilities are far more robust, allowing users to do things like browse across annotation sets, look for nonsyntenic or syntenic orthologs, and explore the regions around a certain region. A user may contribute almost any kind of information to BRAD, including a gene ID from *B. rapa* or *A. thaliana*, a geographic location, or a genetic marker. There are three public databases that provide access to *B. oleracea* genome sequence data: Brassica.info (http://www.brassica.info/resource/databases.php), the AAFC Comparative Genome Viewer (http://brassica.agr.gc.ca/navigation/viewer f.asp), and the *Brassica* Genome Gateway (http://brassica.bbsrc.ac.uk/). These databases include some *B. oleracea* genomic data, such as cloned genes, ESTs, and QTLs. Bolbase (http://ocri-genomics.org/bolbase), a web-based database containing genomic sequences information and genome-wide relevant data, was designed to facilitate the access, search, modeling, and understanding of the *B. oleracea* exome sequencing, annotation, configuration, and evolvement. This collection facilitates study on gene function, comparative genomics, and the evolution of highly associated Brassicaceae species, as well as modern breeding developments within the *Brassica* genus. In addition to 45,758 predicted genes, 13,382 transposable elements, and 3,581 noncoding RNAs, there are nine completely constructed chromosomes and 1,848 scaffolds. Information regarding comparative genomics includes syntenic regions between *B. oleracea*, *Brassica rapa*, and *Arabidopsis thaliana*, synonymous (Ks) and nonsynonymous (Ka)

substitution rates between orthologous gene pairs, gene families, or clusters, and variations in the number, classification, and distribution of mobile genetic elements. Bolbase's search and data mining capabilities include a text search, a local BLAST browser, and a customized GBrowse tool that can be employed to retrieve tags of genome elements, locate related sequences, and show syntenic areas across species. Users may access comprehensive genetic data and conduct comparative genomics research using an interface that is extremely apparent.

12.2 Targeted Databases of Genome, Transcriptome, and Proteome in Brassicas

In addition to sequence analysis and assembling of DNA, research on *B. oleracea* proteomics, transcriptomics, and metabolomics has revealed patterns of gene expression (Wei et al. 2021), metabolite abundance (Parkin et al. 2014; Liu et al. 2014), and protein (Zhao et al. 2020) in a broad range of genotypes. The genomic analysis for just two *B. oleracea* species is insufficient, even if the BRAD V3.0 database has genomic data for a range of vegetable *Brassica* species (Chen et al. 2022). In response to this need, Wang et al. (2022) developed the *B. oleracea* Genome Database (BoGDB), which is the first omic database and a user-friendly platform for doing research on *B. oleracea*. This database includes data on the organism's genome, transcriptome, and metabolome. In order to get the most out of multiomics data, the BoGDB database incorporates a wide variety of useful genomic modules, such as "Gene Search," "Heatmap," "Genome Browser," "Genome," "Tools," "Metabolic," and "Variation." This lays the groundwork for crop genomic, genetic, and molecular design breeding experiments using *B. oleracea*, and it does so in an approachable manner.

13 Social, Political, and Regulatory Concerns

Regulation is essential for guiding the economic growth of any nation because it grants rights on tangible or corporeal property, intangible or incorporeal properties such as intellectual property (IP), exploitation of natural resources, etc. in the form of controlling human or societal behavior with the force of law by means of certain rules or defined norms. To increase productivity and growth, every nation needs improved regulation, primarily to alleviate corporate regulation, to promote job creation, to lower barriers to international commerce and investment, and to bolster investment by implementing a number of financial sector reforms. In order to enhance the uniformity and quality of the regulatory environment, the regulatory framework must be streamlined by addressing specific regulatory challenges across nations and also inside the country, particularly between federal, state, and municipal governments. Throughout the years, as a result of shifting political and societal landscapes, a number of regulations have been revised in conjunction with the development of new technologies to alleviate the difficulties in managing global

regulatory compliances and to meet challenges in compliance and enforcement through a set of transparent, consistent, and nondiscriminatory rules in order to create dynamic and competitive situations.

13.1 Genesis

In 1944, during World War II, the United Nations designated a comprehensive trade barrier that sharply divided nations and wreaked havoc on business and the economy. Thus, a conference known as the Bretten Woods Monetary Conference was convened with 44 member nations to form the World Bank, IMF (International Monetary Fund), and ITO (International Trade Organization). GATT was founded in 1947 with the purpose of regulating trade as part of a wider plan for the post-World War II economy's recovery via the removal of trade subsidies, quantitative constraints, and tariff barriers. The GATT was a contract, not a corporation. The history of the GATT may be divided into three major phases. The first round, which ran from 1947–1950 and was known as the Torquay Round, was largely concerned with deciding which goods would be included by the agreement and freezing present tariff levels. The second phase, which ran three cycles from 1959–1979, focused on reducing tariffs. Only the Uruguay Round, which ran from 1986–1994, comprised the third phase, which substantially enlarged the agreement to encompass more sectors, such as intellectual property, services, capital, and agriculture. The WTO resulted from these negotiations. In 1995, the GATT was superseded by the WTO. India is one of the founding countries to join the World Trade Organization, which was founded by the Marrakesh Agreement, in compliance with the TRIPS agreement.

- Patents must be issued in all "fields of technology," but various exceptions for the public interest may be established (Art. 27.2 and 27.3).
- Article 27 of the TRIPS Agreement outlines the innovations that nations must make accessible for patenting and those that they may prohibit from patenting.
- Patentable inventions may be both goods and processes, and they should normally span all fields of technology.

13.2 Patent and IPR Issues

Intellectual property rights (IPR) are legal rights granted to the inventors or producers for a limited time to prevent others from replicating, producing, utilizing, or selling a protected substance, process, ideas, or innovations. Under Indian law, there are a variety of intellectual properties (IPs), including the Indian Patent Act of 1970, the India Copy Right Act of 1957, the Trade Marks Act of 1999, the Designs Act of 2000, the Geographical Indications of Goods (Registration and Protection) Act of 1999, the semiconductor Integrated Circuits Layout Design Act of 2000, and the Protection of Original Works Act of 1999. The Protection of Plant Varieties and Farmers' Rights Act (PPVFRA) of 2001 is the most recent intellectual property

legislation in India. The Indian Copy Right Act, which was passed in 1957, is the country's oldest IP law.

A patent is a 20-year monopoly that the government provides to an inventor over his or her creation from the date of filing. Patents protect both new and improved versions of existing concepts. An inventor may get a patent for his/her invention, as may anybody he/she assigns the patent to. The inventor or his assignee may get a patent by filing an application to the patent office in accordance with the nation's Patent Act requirements. Only original, obscure, and industrially applicable innovations are eligible for patent protection (utility).

- Novelty (newness): It must possess a characteristic that is not present in the corpus of prior work in the relevant technological domain.
- Nonobviousness (innovative step): It must indicate an imaginative step that a person with an ordinary comprehension of the technical field could not infer.
- Utility (industrial applicability): The invention must be advantageous and applicable to industry.

Section 3 (j) of the Indian Patent Act did not allow for the patenting of agricultural and horticultural production systems. Since 2005, India has been required to comply with the TRIPS Agreement in order to make its domestic laws compliant with TRIPS. In the past, India had no laws to conserve plant types or genetic resources, as there was no need. After India signed TRIPS, Article 27 (3) (b) of TRIPS mandated that member countries guarantee protection for plant types. UPOV places restrictions on the rights of farmers. In light of its socioeconomic circumstances, India established a *sui generis* strategy, combining the rights of breeders, farmers, and researchers and adding FRs to its Act. To improve the efficiency of the *sui generis* system, the PPVFRA was adopted in 2001. Its rules were created in 2003, and the Authority was founded in 2005, not only to offer intellectual property protection for plant types, but also to protect the interests of farmers.

13.3 Access, Benefit Sharing, and Disclosure of Genetic Resources (ABS)

The genetic resources (GRs) and other biological resources are key components of the system of intellectual property (IP). These resources may be obtained from plants, microorganisms, animals, or any other source holding functional units of heredity. Genetic resources derived from humans are excluded from this group. In research and development, the utilization of genetic resources is essential to the creation of a large diversity of commodities and services for the benefit of humanity. Access to GRs is necessary for these activities in order to utilize GRs based on an equal and fair distribution of benefits between users and providers. The Nagoya protocol specifies the GRs in accordance with Article 15(7) of the Convention on Biological Diversity, which states that access and benefit-sharing objectives must be met with regard to genetic resources with greater legal certainty and transparency for

both providers and users of genetic resources. In certain instances, in addition to GRs, indigenous and local community (ILC) traditional knowledge related with GRs is also incorporated. Depending on the aim of usage (research and knowledge) or the product output, the type of the benefit sharing may be monetary or nonmonetary (sharing royalties from a commercial product). In order to secure the equitable and fair exchange of genetic resources (GRs) and their advantages, the provider of access and benefit sharing (ABS) must get prior informed consent (PIC) from a user before signing of mutually agreed terms (MAT).

13.4 Traditional Knowledge

Traditional knowledge (TK) refers to the skills, discoveries, and practices that indigenous and local communities (ILC) have acquired over centuries of experience in respect to GRs. These cultural heritages are comprised of information that has been passed down through the centuries and altered to meet the needs of the local community. Traditional knowledge is important for determining the human benefits of GRs derived from plants, animals, or microorganisms. Information on the beneficial features of these GRs is used in both research and the marketing of items used for a range of essential reasons, including food, medicine, and agricultural practices, among others. In order to protect biodiversity and utilize it in a sustainable way, the Convention on Biological Diversity (CBD) incorporated rules for the use and preservation of indigeneous technical or conventional knowledge. There are three primary safeguarding measures for traditional knowledge. The first emphasizes the significance of maintaining traditional knowledge as a type of cultural heritage. The second section examines the safeguarding of traditional knowledge as a collective human right. Thirdly, the WTO and WIPO investigate the application of existing or new special measures to safeguard traditional knowledge. The result is the "Biodiversity Act of 2002."

13.5 Farmer Rights (FRs)

Both the Convention on Biological Diversity (1992) and the International Treaty on Plant Genetic Resources for Food and Agriculture (2001) lay a strong focus in Article 9 of their separate conventions on the role of rural and indigenous societies in preserving and developing biodiversity. The most sophisticated law of India is the only one that gives farmers a wide variety of rights, including the capacity to register their own kinds. The rights that farmers have under the PPVFR Act of 2001 are described here.

13.5.1 Seed Rights of Farmers

Farmers still have the same rights to their agricultural goods as they did before to the Act's implementation, including the freedom to keep, use, sow, resow, trade, share,

and sell any seeds of kinds covered by the Act. Farmers do, however, have the right to sell unbranded seed of a variety that is covered by this Act.

13.5.2 Farmers' Rights to Register Traditional Varieties

Farmers have the right to register any new varieties they independently developed as well as any old kinds they generated or protected. It costs nothing for farmers to register their varieties.

13.5.3 Farmers' Rights to Reward and Recognition

The Act includes provisions to reward and recognize individual farmers, agricultural communities, and tribal groups for their accomplishments to the conservation of agricultural plant diversity and the enhancement of plant genetic resources (PGR) via the Plant Genome Savior Award.

13.5.4 Farmers' Rights for Benefit Sharing

Under the Act, farmers or tribal groups that supplied the kinds utilized as parents are entitled to an equal share of the benefits from the new variety. When their variety is utilized as a parent, farmers have the option to make claims for benefit sharing. The Act mandates that the registered seed be marketed with a statement on varietal agronomic performance and commercial features. Farmers have the right to get compensation for losses caused by the registered variety.

13.5.5 Farmers' Rights to Get Compensation for Losses Suffered

The PBR holder is required to provide restitution to impacted farmers in the event that farmers are unable to deliver the promised performance.

13.5.6 Farmers' Rights to Reimbursement for Covert Use of Traditional Varieties

When submitting a registration application, applicants are obliged to identify the traditional varieties that were utilized as parents to produce the new variety. In the case that the applicant conceals information about the community's use of traditional varieties or knowledge and if this is done, a third party may file claims for compensation, which the Authority will then decide to provide. Farmers have the right to purchase seeds from registered varieties at a fair price.

13.5.7 Farmers' Rights for the Seeds of Registered Variety

Farmers are permitted to acquire seeds from recognized varieties. If this condition is not met, the breeder's exclusive ownership of the variety is suspended under the compulsory licensing regulation, and the breeder is compelled to grant a licence to a competent legal organization for the production, distribution, and sale of the variety's seeds.

13.5.8 Farmers' Rights to Receive Free Services

Farmers are exempt from all fees, including those associated with the registration of varieties, conducting DUS tests on varieties, renewing registration, fees associated

with any applications, fees associated with legal proceedings, etc., in consideration of their limited financial resources and to facilitate their access to the Act's benefits.

13.5.9 Farmers' Rights to Be Protected Against Innocent Infringement

Legal precedent dictates that inadvertent lawbreaking is not regarded a valid defense. However, there are protections in place to avoid the violation of the rights of innocent persons. Act violations committed by farmers cannot be tried in court.

13.6 Treaties and Conventions

13.6.1 Convention on Biological Diversity (CBD)

It is a 1992 UNCED international agreement that took effect on December 29, 1993. Its key goals are the maintenance of biological variability (biodiversity), the sustainable use of resources, and the equitable and fair sharing of the value gained from the use of genetic traditional knowledge and resources.

13.6.2 The Nagoya Protocol on Access and Benefit Sharing

It is a supplement to the Convention on Biological Diversity (CBD) from 1992. It was approved in Nagoya, Japan, in 2010 and entered into force on October 12, 2014; India is a signatory. This protocol includes genetic resource access, profit sharing, and compliance processes as member responsibilities.

13.6.3 International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA)

It is an all-encompassing accord consistent with the CBD. The treaty was adopted during the 2001 FAO meeting and entered into force on June 29, 2004. The pact tackles conservation, utilization, international cooperation, technical aid, and the rights of farmers. It also establishes regulations for access and benefit sharing for selected ex situ and in situ crops (35 food crops and 29 forage species) while protecting property rights.

13.6.4 The WTO's Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPs)

The World Trade Organization (WTO) is responsible for enforcing the TRIPS International Agreement, which establishes the fundamental standards for various forms of intellectual property (IP) protection. The Uruguay round of the General Agreement on Tariffs and Trade (GATT) concluded in 1994; this was the final round of negotiations. It stipulates that member nations protect plant species through effective sui generis (standalone) regimes, patents, or a combination of the two. The majority of countries have laws that adhere to the UPOV Convention and meet this requirement. The UPOV Council rejected a law protecting the rights of Indian plant breeders because it violated the conditions of the treaty.

13.6.5 The International Union for the Protection of New Varieties of Plants (UPOV)

It is a multinational organization headquartered in Geneva, Switzerland. It was developed by the International Convention for the Protection of New Varieties of Plants with the objective to optimize plant breeding efforts by offering a paradigm for assuring protection for plant breeder's rights under UPOV.

13.7 Participatory Plant Breeding (PPB)

Participatory breeding is a collective method of plant genetic improvement used by farmers to enhance crops and create new plant varieties in collaboration with scientists, consumers, and other stakeholders with breeding objectives based on their particular region, environment, culture, and need, among other factors. Promising lines are chosen, varieties are generated, and evaluated and novel procedures are used for further multiplication, assessment, seed production, and distribution to farmers using the local germplasm or GRs. The goal of participatory breeding is to guarantee that farmers have access to genetic lines so they may get varieties that meet their needs for quality and livelihood usage. This strategy promotes genetic diversity. In order to effectively employ farmers' varieties (traditional varieties/landraces/wild relatives) and GRs to generate superior cultivars, participatory breeding programs are being developed. Farmers are also offered a compensation and support system for their contribution to the global genetic pool.

14 Future Perspectives

Throughout the course of the past several decades, there has been a consistent rise in the degree of knowledge about the significance of glucosinolates, both in terms of their involvement in the process of plant self-defense and in terms of their importance for human and animal nutrition. As a direct consequence of this discovery, a significant amount of research into the genetic underpinnings of glucosinolate production has been carried out. However, the QTL analysis showed that there are additional genetic traits in *Brassica* crops that need further investigation. These characteristics include: The most important genes in these plants have been thoroughly studied. Other facets, such as the method in which glucosinolate genes and systems interact with environmental stressors, are examples of areas that are examples of topics that are obtaining an increasing relevance in current society. In light of the fact that other species of the Brassicaceae that are not in the genus *Brassica* are recognized as being good sources of beneficial chemicals, it is feasible to include them into one's diet as a result of the data provided here. These animals each have a unique glucosinolate pattern, which is distinguishable from the others. Investigating the genetics behind these traits will provide exciting insights that might be used in the breeding of other Brassicaceae species. Breeding efforts are now integrating a complete understanding of the genetics of glucosinolates with the mechanism by

which they are inherited in order to change the levels of these useful metabolites. This is being done in an attempt to modify the levels of these helpful metabolites. This process will be sped up as a consequence of the sequencing of the primary *Brassica* crops and the commonalities that now exist between those crops and *Arabidopsis*.

References

Abd-Elsalam KA (2020) Multifunctional hybrid nanomaterials for sustainable agri-food and ecosystems: a note from the editor. Elsevier Inc., Amsterdam

Agerbirk N, Olsen CE (2012) Glucosinolate structures in evolution. Phytochemistry 77:16–45

Ahmadiani N, Robbins RJ, Collins TM, Giusti MM (2014) Anthocyanins contents, profiles, and color characteristics of red cabbage extracts from different cultivars and maturity stages. J Agric Food Chem 62:7524–7531

Anilakumar KR, Khanum F, Bawa AS (2006) Dietary role of glucosinolate derivatives: a review. J Food Sci Technol 43:8–17

Araki R, Hasumi A, Nishizawa OI, Sasaki K, Kuwahara A, Sawada Y, Totoki Y, Toyoda A, Sakaki Y, Li Y et al (2013) Novel bioresources for studies of *Brassica oleracea*: identification of a kale MYB transcription factor responsible for glucosinolate production. Plant Biotechnol J 11:1017–1027

Armah CN, Traka MH, Dainty JR, Defernez M, Janssens A, Leung W, Doleman JF, Potter JF, Mithen RF (2013) A diet rich in high-glucoraphanin broccoli interacts with genotype to reduce discordance in plasma metabolite profiles by modulating mitochondrial function. Am J Clin Nut 98:712–722

Artemyeva AM, Solovieva AE, Kocherina NV, Chesnokov YV (2014) QTL analysis of biochemical traits of quality in *Brassica rapa* L. Veg Crop Russ 1:10–13

Assefa AD, Choi S, Lee JE, Sung JS, Hur OS, Ro NY, Lee HS, Noh J, Hwang A, Ko HC, Chung YJ (2019) Determination of glucosinolate contents in Brassica germplasm collections and inter-& intra-leaves distribution pattern using UPLC-MS/MS multiple reaction monitoring scan mode. BioRxiv:569889

Augustine R, Majee M, Gershenzon J, Bisht NC (2013) Four genes encoding MYB28, a major transcriptional regulator of the aliphatic glucosinolate pathway, are differentially expressed in the allopolyploid *Brassica juncea*. J Exp Bot 64:4907–4921

Baenas N, García-Viguera C, Moreno DA (2014) Biotic elicitors effectively increase the glucosinolates content in Brassicaceae sprouts. J Agric Food Chem 62:1881–1889

Bagheri H, El-Soda M, Kim HK, Fritsche S, Jung C, Aarts MGM (2013) Genetic analysis of health-related secondary metabolites in a *Brassica rapa* Recombinant inbred line population. Int J Mol Sci 14:15561–15577

Baik HY, Juvik JA, Jeffery EH, Wallig MA, Kushad M, Klein BP (2003) Relating glucosinolate content and flavor of broccoli cultivars. J Food Sci 68:1043–1050

Baskar V, Park SW (2015) Molecular characterization of BrMYB28 and BrMYB29 paralogous transcription factors involved in the regulation of aliphatic glucosinolate profiles in *Brassica rapa* ssp. *pekinensis*. Crit Rev Biol 338:434–442

Belser CB, Istace E, Denis M, Dubarry F, Baurens C, Falentin M et al (2018) Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. Nat Plants 4:879–887

Bennett RN, Mellon FA, Rosa EAS, Perkins L, Kroon PA (2004) Profiling glucosinolates, flavonoids, alkaloids, and other secondary metabolites in tissues of *Azima tetracantha* L. (Salvadoraceae). J Agric Food Chem 52:5856–5862

Berhow MA, Polat U, Glinski JA, Glensk M, Vaughn SF, Isbell T, Ayala-Diaz I, Marek L, Gardner C (2013) Optimized analysis and quantification of glucosinolates from *Camelina sativa* seeds by reverse-phase liquid chromatography. Ind Crop Prod 43:119–125

Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci 48:1649–1664

Bhandari SR, Jo JS, Lee JG (2015) Comparison of glucosinolate profiles in different tissues of nine *Brassica* crops. Molecules 20:15827–15841

Block G, Patterson B, Subar A (1992) Fruit, vegetables, and cancer prevention: a review of the epidemiological evidence. Nutr Cancer 18:1–29

Bodnaryk RP (1992) The international journal of plant biochemistry effects of wounding on glucosinolates in the cotyledons of oilseed rape and mustard. Phytochemistry 31:2671–2677

Bones AM, Rossiter JT (1996) The myrosinase-glucosinolate system, its organisation and biochemistry. Physiol Planta 97:194–208

Bongoni R, Verkerk R, Steenbekkers B, Dekker M, Stieger M (2014) Evaluation of different cooking conditions on broccoli (*Brassica oleracea* var. *italica*) to improve the nutritional value and consumer acceptance. Plant Foods Hum Nutr 69:228–234

Brader G, Tas É, Palva ET (2001) Jasmonate-dependent induction of indole glucosinolates in Arabidopsis by culture filtrates of the nonspecific pathogen *Erwinia carotovora*. Plant Physiol 126:849–860

Bradshaw J, Wilson R (2012) Kale population improvement and cultivar production. Euphytica 184:275–288

Bramley PM (2002) Regulation of carotenoid formation during tomato fruit ripening and development. J Exp Bot 53:2107–2113

Branca F, Li G, Goyal S, Quiros CF (2002) Survey of aliphatic glucosinolates in Sicilian wild and cultivated Brassicaceae. Phytochemistry 59:717–724

Branca F, Chiarenza GL, Cavallaro C, Gu H, Zhao Z, Tribulato A (2018) Diversity of Sicilian broccoli (*Brassica oleracea* var. *italica*) and cauliflower (*Brassica oleracea* var. *botrytis*) landraces and their distinctive bio-morphological, antioxidant, and genetic traits. Genet Resour Crop Evol 65:485–502

Britton G (1995) Structure and properties of carotenoids in relation to function. FASEB J 9: 1551–1558

Broun P (2005) Transcriptional control of flavonoid biosynthesis: a complex network of conserved regulators involved in multiple aspects of differentiation in Arabidopsis. Curr Opin Plant Biol 8: 272–279

Brown AF, Yousef GG, Chebrolu KK, Byrd RW, Everhart KW, Thomas A, Reid R, Parkin IAP, Sharpe AG, Oliver R et al (2014) High-density single nucleotide polymorphism (SNP) array mapping in *Brassica oleracea*: identification of QTL associated with carotenoid variation in broccoli florets. Theor Appl Genet 127:2051–2064

Brown AF, Yousef GG, Reid RW, Chebrolu KK, Thomas A, Krueger C, Jeffery E, Jackson E, Juvik JA (2015) Genetic analysis of glucosinolate variability in broccoli florets using genome anchored single nucleotide polymorphisms. Theor Appl Genet 128:1431–1447

Brüsewitz G, Cameron BD, Chasseaud LF, Görler K, Hawkins DR, Koch H et al (1977) The metabolism of benzyl isothiocyanate and its cysteine conjugate. Biochem J 162:99–107

Burdzinski C, Wendell DL (2007) Mapping the anthocyaninless (*anl*) locus in rapid-cycling *Brassica rapa* (RBr) to linkage group R9. BMC Genet 8:6456–6468

Cai X, Wu J, Liang J, Lin R, Zhang K, Cheng F et al (2020) Improved *Brassica oleracea* JZS assembly reveals significant changing of LTR-RT dynamics in different morphotypes. Theor Appl Genet 133:3187–3199

Cao H, Liu R, Zhang J, Liu Z, Fan S, Yang G, Jin Z, Pei Y (2021) Improving sulforaphane content in transgenic broccoli plants by overexpressing MAM1, FMOGS–OX2, and Myrosinase. Plant Cell Tissue Organ Cult 46:461–471

Carlson DG, Daxenbichler ME, Tookey HL (1987a) Glucosinolates in turnip tops and roots: cultivars grown for greens and/or roots. J Am Soc Hortic Sci 112:179–183

Carlson DG, Daxenbichler ME, VanEtten CH, Kwolek WF, Williams PH (1987b) Glucosinolates in crucifer vegetables: broccoli, Brussels sprouts, cauliflower, collards, kale, mustard greens, and kohlrabi. J Am Soc Hortic Sci 112:173–178

Cartea ME, Velasco P (2008) Glucosinolates in Brassica foods: bioavailability in food and significance for human health. Phytochem Rev 7:213–229
Cartea ME, Rodríguez VM, de Haro A, Velasco P, Ordás A (2008a) Variation of glucosinolates and nutritional value in nabicol (*Brassica napus pabularia* group). Euphytica 159:111–122
Cartea ME, Velasco P, Obregón S, Padilla G, de Haro A (2008b) Seasonal variation in glucosinolate content in *Brassica oleracea* crops grown in northwestern Spain. Phytochemistry 69:403–410
Castro-Acosta ML et al (2016) Berries and anthocyanins: promising functional food ingredients with postprandial glycaemia-lowering effects. Proc Nutr Soc 75:342–355
Cazzonelli CI (2011) Carotenoids in nature: insights from plants and beyond. Funct Plant Biol 38: 833–847
Celenza JL, Quiel JA, Smolen GA, Merrikh H, Silvestro AR, Normanly J, Bender J (2005) The *Arabidopsis* ATR1 MYB transcription factor controls indolic glucosinolate homeostasis. Plant Physiol 137:253–262
Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B et al (2014) Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. Science 345:950–953
Charron CS, Clevidence BA, Britz SJ, Novotny JA (2007) Effect of dose size on bioavailability of acylated and nonacylated anthocyanins from red cabbage (*Brassica oleracea* L. var. *capitata*). J Agric Food Chem 55:5354–5362
Chaudhary A, Choudhary S, Sharma U, Vig AP, Singh B, Arora S (2018) Purple head broccoli (*Brassica oleracea* L. var. *italica* Plenck), a functional food crop for antioxidant and anticancer potential. J Food Sci Technol 55:1806–1815
Chayut N, Yuan H, Ohali S, Meir A, Sa'ar U et al (2017) Distinct mechanisms of the ORANGE protein in controlling carotenoid flux. Plant Physiol 173:376–389
Chen C, Xiao L, Li X, Du D (2018a) Comparative mapping combined with map-based cloning of the *Brassica juncea* genome reveals a candidate gene for multilocular rapeseed. Front Plant Sci 9:1744
Chen F, Dong G, Ma X, Wang F, Zhang Y, Xiong E, Wu J, Wang H, Qian Q, Wu L, Yu Y (2018b) UMP kinase activity is involved in proper chloroplast development in rice. Photosynth Res 137:53–67
Chen H, Wang T, He X, Cai X, Lin R, Liang J et al (2022) BRAD V3.0: an upgraded Brassicaceae database. Nucleic Acids Res 50:D1432–D1441
Cheng F, Liu S, Wu J et al (2011) BRAD, the genetics and genomics database for Brassica plants. BMC Plant Biol 11:136
Chi W, He B, Mao J, Li Q, Ma J, Ji D, Zou M, Zhang L (2012) The function of *RH22*, a DEAD RNA helicase, in the biogenesis of the 50S ribosomal subunits of *Arabidopsis* chloroplasts. Plant Physiol 158:693–707
Chiu LW, Zhou X, Burke S, Wu X, Prior RL, Li L (2010) The purple cauliflower arises from activation of a MYB transcription factor. Plant Physiol 154:1470–1480
Ciska E, Martyniak-Przybyszewska B, Kozlowska H (2000) Content of glucosinolates in cruciferous vegetables grown at the same site for two years under different climatic conditions. J Agric Food Chem 48:2862–2867
Clarke DB (2010) Glucosinolates, structures and analysis in food. Analyt Methods 2:310–325
Connolly EL, Sim M, Travica N, Marx W, Beasy G, Lynch GS, Bondonno CP, Lewis JR, Hodgson JM, Blekkenhorst LC (2021) Glucosinolates from cruciferous vegetables and their potential role in chronic disease: investigating the preclinical and clinical evidence. Front Pharmacol:2964
Corella D, Carrasco P, Sorlí JV, Estruch R, Rico-Sanz J, Martínez-González MÁ et al (2013) Mediterranean diet reduces the adverse effect of the *TCF7L2-rs7903146* polymorphism on cardiovascular risk factors and stroke incidence: a randomized controlled trial in a high-cardiovascular-risk population. Diabet Care 36:3803–3811
Cornelis MC, El-Sohemy A, Campos H (2007) *GSTT1* genotype modifies the association between cruciferous vegetable intake and the risk of myocardial infarction. Am J Clin Nutr 86:752–758
Cotton SC, Sharp L, Little J, Brockton N (2000) Glutathione S-transferase polymorphisms and colorectal cancer: a HuGE review. Am J Epidemiol 151:7–32

Crisp P, Angell SM (1985) Genetic control of green colour in cauliflower. Ann Appl Biol 107:601
Crisp P, Tapsell CR (1992) Cauliflower *Brassica oleracea* L. In: Kalloo G, Berg BO (eds) Genetic improvement of vegetable crops. Pergamon Press, Oxford, pp 157–178
Crisp PWDG, Walkey DGA, Bellman E, Roberts E (1975) A mutation affecting curd colour in cauliflower (*Brassica oleracea* L. var. *botrytis* DC). Euphytica 24:173–176
Cunningham FX, Gantt E (1998) Genes and enzymes of carotenoid biosynthesis in plants. Annu Rev Plant Physiol Plant Mol Biol 49:557–583
Cunningham FX, Gantt E (2002) Molecular control of floral pigmentation: carotenoids. In: Vainstein A (ed) Breeding for ornamentals: classical and molecular approaches. Kluwer Academic Publishers, Boston, pp 273–293
Cunningham FJ, Goh NS, Demirer GS, Matos JL (2018) Nanoparticle-mediated delivery towards advancing plant genetic engineering. Trends Biotechnol 36:882–897
Del Carpio DP, Basnet RK, Arends D, Lin K, De Vos RCH, Muth D, Kodde J, Boutilier K, Bucher J, Wang XW et al (2014) Regulatory network of secondary metabolism in *Brassica rapa*: insight into the glucosinolate pathway. PLoS One 9:e107123
Delourme R, Falentin C, Huteau V, Clouet V, Horvais R, Gandon B et al (2006) Genetic control of oil content in oilseed rape (*Brassica napus* L.). Theor Appl Genet 113:1331–1345
Demirer GS, Silva TN, Jackson CT, Thomas JB, Ehrhardt DW, Rhee SY et al (2021) Nanotechnology to advance CRISPR–Cas genetic engineering of plants. Nat Nanotechnol 12:1–8
Dickson MH, Wallace DH (1986) Cabbage breeding. In: Bassett MJ (ed) Breeding vegetable crops. AVI Publishing Co., Westport, CT, pp 395–432
Dickson MH, Lee CY, Blamble AE (1988) Orange-curd high carotene cauliflower inbreds, NY 156, NY 163, and NY 165. Hortic Sci 23:778–779
Ding Y, Jian Y (2010) An orange cauliflower hybrid 'Jinyu 60' with high β-carotene. Acta Hortic 856:261
Dosz EB, Jeffery EH (2013) Commercially produced frozen broccoli lacks the ability to form sulforaphane. J Funct Foods 5:987–990
Doughty KJ, Kiddle GA, Pye BJ, Wallsgrove RM, Pickett JA (1995) Selective induction of glucosinolates in oilseed rape leaves by methyl jasmonate. Phytochemistry 38:347–350
Egner PA, Chen JG, Zarth AT, Ng DK, Wang JB, Kensler KH, Jacobson LP, Muñoz A, Johnson JL, Groopman JD, Fahey JW (2014) Rapid and sustainable detoxication of airborne pollutants by Broccoli Sprout beverage: results of a randomized clinical trial in China detoxication of air pollutants with Broccoli Sprout beverage. Cancer Prevent Res 7:813–823
Egorova KV, Sinyavina NG, Artemyeva AM, Kocherina NV, Chesnokov YV (2021) QTL analysis of the content of some bioactive compounds in *Brassica rapa* L. grown under light culture conditions. Horticulturae 7:583
El-Mogy MM, Mahmoud AWM, El-Sawy MBY, Parmar A (2019) Pre-harvest foliar application of mineral nutrients to retard chlorophyll degradation and preserve bio-active compounds in broccoli. Agronomy 9:711
Ettlinger MG, Kjaer A (1968) Sulfur compounds in plants. Recent Adv Phytochem 1:144
Ettlinger MG, Lundeen AJ (1956) The structures of sinigrin and sinalbin; an enzymatic rearrangement. J Am Chem Soc 78:4172–4173
Fahey JW, Zhang Y, Talalay P (1997) Broccoli sprouts: an exceptionally rich source of inducers of enzymes that protect against chemical carcinogens. Proc Natl Acad Sci U S A 94:10367–10372
Fahey JW, Zalcmann AT, Talalay P (2001) The chemical diversity and distribution of glucosinolates and isothiocyanates among plants. Phytochemistry 56:5–51
Fahey JW, Haristoy X, Dolan PM, Kensler TW, Scholtus I (2002) Sulforaphane inhibits extracellular, intracellular, and antibiotic-resistant strains of *Helicobacter pylori* and prevents benzo[a] pyrene-induced stomach tumors. Proc Natl Acad Sci U S A 99:7610–7615
Fang H, Xu J, Ding D, Jackson SA, Patel IR et al (2010) An FDA bioinformatics tool for microbial genomics research on molecular characterization of bacterial foodborne pathogens using microarrays. BMC Bioinformatics 11:S4
FAOSTAT (2020) Food and Agriculture Organization, Rome. Online accessed on 28 August 2022

Farnham MW, Kopsell DA (2009) Importance of genotype on carotenoid and chlorophyll levels in broccoli heads. HortScience 44:1248–1253
Farnham MW, Wilson PE, Stephenson KK, Fahey JW (2004) Genetic and environmental effects on glucosinolate content and chemoprotective potency of broccoli. Plant Breed 123:60–65
Faulkner K, Mithen R, Williamson G (1998) Selective increase of the potential anticarcinogen 4-methylsulphinylbutyl glucosinolate in broccoli. Carcinogenesis 19:605–609
Fernandes I, Marques C, Évora A, Faria A, Calhau C, Mateus N, de Freitas V (2019) Anthocyanins: nutrition and health. In: Mérillon J-M, Ramawat KG (eds) Bioactive molecules in food. Springer, Cham, pp 1097–1133
Fiers MW, Kleter GA, Nijland H, Peijnenburg AA, Nap JP et al (2004) Allermatch™, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. BMC Bioinformatics 5:133
Fowke JH, Chung FL, Jin F, Qi D, Cai Q, Conaway C, Cheng JR, Shu XO, Gao YT, Zheng W (2003) Urinary isothiocyanate levels, brassica, and human breast cancer. Cancer Res 63:3980–3986
Francisco M, Ali MAA, Ferreres F, Moreno-Fernández D, Ángel Velasco P, Soengas P (2016) Organ-specific quantitative genetics and candidate genes of phenylpropanoid metabolism in *Brassica oleracea*. Front Plant Sci 6:1240
Frerigmann H, Gigolashvili T (2014) MYB34, MYB51, and MYB122 distinctly regulate indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. Mol Plant 7:814–828
Gad MA, Li M, Ahmed FK, Almoammar H (2020) Nanomaterials for gene delivery and editing in plants: challenges and future perspective. Elsevier Inc., Amsterdam
Gao M, Li G, Yang B, McCombie WR, Quiros CF (2004) Comparative analysis of a Brassica BAC clone containing several major aliphatic glucosinolate genes with its corresponding Arabidopsis sequence. Genome 47:666–679
Gao J, Yu X, Ma F, Li J (2014) RNA-Seq analysis of transcriptome and glucosinolate metabolism in seeds and sprouts of broccoli (*Brassica oleracea* var. *italica*). PLoS One 9:e88804
Gao H, Yang X, Wang H, Qiu N, Chen Y, Wang F, Zhang Y, Li H, Li J, Gao J (2022) Construction of an intragenic SSR-based linkage map and QTL mapping for agronomic traits in chinese cabbage (*Brassica rapa* L. ssp. pekinensis). Horticulturae 8:165
Gardner AM, Brown AF, Juvik JA (2016) QTL analysis for the identification of candidate genes controlling phenolic compound accumulation in broccoli (*Brassica oleracea* L. var. *italica*). Mol Breed 36:81
Garte S, Gaspar L, Alexandrie AK, Ambrosone C, Autrup H, Autrup JL (2001) Metabolic gene polymorphism frequencies in control populations. Cancer Epidemiol Biomark Prev 10:1239–1249
Gasper AV, Al-Janobi A, Smith JA, Bacon JR, Fortun P, Atherton C, Taylor MA, Hawkey CJ, Barrett DA, Mithen RF (2005) Glutathione S-transferase M1 polymorphism and metabolism of sulforaphane from standard and high-glucosinolate broccoli. Am J Clin Nutr 82:1283–1291
Geu Flores F, Olsen C, Halkier BA (2009) Towards engineering glucosinolates into non-cruciferous plants. Planta 229:261–270
Giamoustaris A, Mithen RF (1996) Genetics of aliphatic glucosinolates. IV. Side-chain modification in *Brassica oleracea*. Theor Appl Genet 93:1006–1010
Gigolashvili T, Berger B, Flugge U-I (2009) Specific and coordinated control of indolic and aliphatic glucosinolate biosynthesis by R2R3-MYB transcription factors in *Arabidopsis thaliana*. Phytochem Rev 8:3–13
Giovannucci E, Rimm EB, Liu Y, Stampfer MJ, Willett WC (2003) A prospective study of cruciferous vegetables and prostate cancer. Cancer Epidemiol Biomark Prev 12:1403–1409
Gonzalez A, Zhao M, Leavitt JM, Lloyd AM (2008) Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. Plant J 53:814–827
Gray AR (1992) Broccoli. *Brassica oleracea* L. (Italica group). In: Kalloo G, Berg BO (eds) Genetic improvement of vegetable crops. Pergamon Press, Oxford, pp 61–86
Grotewold E, Sainz MB, Tagliani L, Hernandez JM, Bowen B, Chandler VL (2000) Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. Proc Natl Acad Sci U S A 97

Grubb CD, Abel S (2006) Glucosinolate metabolism and its control. Trends Plant Sci 11:89–100
Grubb DC, Zipp BJ, Ludwig-Muller J, Masuno MN, Molinski TF, Abel S (2004) Arabidopsis glucosyltransferase UGT74B1 functions in glucosinolate biosynthesis and auxin homeostasis. Plant J 40:893–908
Guasch-Ferré M, Bhupathiraju SN, Hu FB (2018) Use of metabolomics in improving assessment of dietary intake. Clin Chem 64:82–98
Guo N, Cheng F, Wu J, Liu B, Zheng S, Liang J, Wang X (2014) Anthocyanin biosynthetic genes in *Brassica rapa*. BMC Genomics 15:1–11
Guo N, Wu J, Zheng SN, Cheng F, Liu B, Liang JL, Cui Y, Wang XW (2015) Anthocyanin profile characterization and quantitative trait locus mapping in Zicaitai (*Brassica rapa* L. ssp. *Chinensis* var. *purpurea*). Mol Breed 35:1–11
Guo N, Wang S, Gao L, Liu Y, Wang X, Lai E et al (2021) Genome sequencing sheds light on the contribution of structural variants to *Brassica oleracea* diversification. BMC Biol 19:93
Hafeez B, Khanif Y, Saleem M (2013) Role of zinc in plant nutrition – a review. Am J Exp Agric 3:374
Halkier BA (2016) General introduction to glucosinolates. Adv Bot Res 80:1–14
Halkier BA, Du L (1997) The biosynthesis of glucosinolates. Trends Plant Sci 2(11):425–431
Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. Annu Rev Plant Biol 57:303–333
Hansen BG, Kliebenstein DJ, Halkier BA (2007) Identification of a flavin-monooxygenase as the S-oxygenating enzyme in aliphatic glucosinolate biosynthesis in *Arabidopsis*. Plant J 50: 902–910
Hansson LE, Nyren O, Bergstro R (1993) Diet and risk of gastric cancer. Int J Cancer 55:181–189
Hayes J, Kelleher M, Eggleston I (2008) The cancer chemopreventive actions of phytochemicals derived from glucosinolates. Eur J Nutr 47:73–88
He Y, Yang Z, Tang M, Yang QY, Zhang Y, Liu S (2022) Enhancing canola breeding by editing a glucosinolate transporter gene lacking natural variation. Plant Physiol 188:1848–1851
Hemlata Singh J, Bhardwaj A, Kumar A, Singh G, Priya K, Giri SK (2022) Comparative frequency distribution of glutathione S-transferase mu (GSTM1) and theta (GSTT1) allelic forms in Himachal Pradesh population. Egyp J Med Hum Genet 23:1–7
Hirani AH, Zelmer CD, McVetty PB, Daayf F, Li G (2013) Homoeologous GSL-ELONG gene replacement for manipulation of aliphatic glucosinolates in *Brassica rapa* L. by marker assisted selection. Front Plant Sci 4:55
Hirschberg J (2001) Carotenoid biosynthesis in flowering plants. Curr Opin Plant Biol 4:210–218
Hirschmann F, Papenbrock J (2015) The fusion of genomes leads to more options: a comparative investigation on the desulfo-glucosinolate sulfotransferases of *Brassica napus* and homologous proteins of *Arabidopsis thaliana*. Plant Physiol Biochem 91:10–19
Hu X, Sullivan-Gilbert M, Gupta M, Thompson SA (2006) Mapping of the loci controlling oleic and linolenic acid contents and development of *fad2* and *fad3* allele-specific markers in canola (*Brassica napus* L.). Theor Appl Genet 113:497–507
Huang H, Cui T, Zhang L, Yang Q, Yang Y, Xie K, Fan C, Zhou Y (2020) Modifications of fatty acid profile through targeted mutation at *BnaFAD2* gene with CRISPR/Cas9-mediated gene editing in *Brassica napus*. Theor Appl Genet 133:2401–2411
Huggins C, Grand LC, Brillantes FP (1961) Mammary cancer induced by a single feeding of polynuclear hydrocarbons, and its suppression. Nature 189:204–207
Jahangir M, Kim HK, Choi YH, Verpoorte R (2009) Health-affecting compounds in Brassicaceae. Comp Rev Food Sci Food Saf 8:31–43
Jan SA, Shinwari ZK, Rabbani MA (2016) Determining genetic divergence among *Brassica rapa* ecotypes through electrophoretic mobility of total seed proteins. J Anim Plant Sci 26:1758–1764
Jenkins JA, Griffiths-Jones S, Shewry PR, Breiteneder H, Mills EC (2005) Structural relatedness of plant food allergens with specific reference to cross-reactive allergens: an in silico analysis. J Allergy Clin Immunol 115:163–170
Jha SK, Singh HR, Prakash P (2017) Dietary fiber and human health: an introduction. In: Dietary fiber for the prevention of cardiovascular disease. Academic Press, pp 1–22
Juge N, Mithen RF, Traka M (2007) Molecular basis for chemoprevention by sulforaphane: a comprehensive review. Cell Mol Life Sci 64:1105–1127

Kaiser AE, Baniasadi M, Giansiracusa D, Giansiracusa M, Garcia M, Fryda Z, Wong TL, Bishayee A (2021) Sulforaphane: a broccoli bioactive phytocompound with cancer preventive potential. Cancers 13:4796

Kalia P (2009) Genetic improvement in vegetable Cruficers. In: Gupta SK (ed) Biology and breeding of Crucifers. CRC Press, Taylor and Francis, Boca Raton, pp 309–342

Kalia P, Muthukumar P, Soi S, Shilpi (2018) Marker-assisted introgression of the *Or* gene for enhancing βcarotene content in Indian cauliflower. Acta Hortic 1203:121–128

Kalia P, Muthukumar P, Sharma R, Asha Devi R, Sharma K, Saha S (2020) *Or* gene led enrichment of β-carotene in Indian cauliflower to tackle malnutrition. Acta Hortic 1282:247–254

Karunarathna NL, Wang H, Harloff HJ, Jiang L, Jung C (2020) Elevatingseed oil content in a polyploid crop by induced mutations in SEEDFATTY ACID REDUCER genes. Plant Biotechnol J 18:2251–2266

Ken H, Matsumoto S, Tsukazaki H, Kondo T, Kubo N, Hirai M (2010) Mapping of a novel locus regulating anthocyanin pigmentation in *Brassica rapa*. Breed Sci 60:76–80

Kensler TW, Ng D, Carmella SG, Chen M, Jacobson LP, Munoz A, Egner PA, Chen JG, Qian GS, Chen TY, Fahey JW (2012) Modulation of the metabolism of airborne pollutants by glucoraphanin-rich and sulforaphane-rich broccoli sprout beverages in Qidong, China. Carcinogenesis 33:101–107

Kim J (2021) Pre-clinical neuroprotective evidences and plausible mechanisms of sulforaphane in Alzheimer's Disease. Int J Mol Sci 22:2929

Kim YC, Cha A, Hussain M, Lee K, Lee S (2020) Impact of Agrobacterium-infiltration and transient overexpression of BroMYB28 on glucoraphanin biosynthesis in broccoli leaves. Plant Biotechnol Rep 14:373–380

Klein M, Reichelt M, Gershenzon J, Papenbrock J (2006) The three desulfoglucosinolate sulfotransferase proteins in *Arabidopsis* have different substrate specificities and are differentially expressed. FEBS J 273:122–136

Kliebenstein DJ (2009) A quantitative genetics and ecological model system: understanding the aliphatic glucosinolate biosynthetic network via QTLs. Phytochem Rev 8:243–254

Klopsch R, Witzel K, Artemyeva A, Ruppel S, Hanschen FS (2018) Genotypic variation of glucosinolates and their breakdown products in leaves of *Brassica rapa*. J Agric Food Chem 66:5481–5490

Kopsell DA, Kopsell DE, Lefsrud MG, Curran-Celentano J, Dukach LE (2004) Variation in lutein, β-carotene, and chlorophyll concentrations among *Brassica oleracea* cultigens and seasons. HortScience 39:361–364

Kopsell DA, McElroy JS, Sams CE, Kopsell DE (2007) Genetic variation in carotenoid concentrations among diploid and amphidiploid rapid-cycling *Brassica* species. HortScience 42:461–465

Koritsas VM, Lewis JA, Fenwick GR (1991) Glucosinolate responses of oilseed rape, mustard and kale to mechanical wounding and infestation by cabbage stem flea beetle (*Psylliodes chrysocephala*). Ann Appl Biol 118:209–221

Kristofferson KB (1924) Contributions to the genetics of *Brassica oleracea*. Hereditas 5:297–364

Kriti A, Ghatak A, Mandal N (2020) Antimycotic efficacy of zinc nanoparticle on dark spore forming phytopathogenic fungi. J Pharmacog Phytochem 9:750–754

Kumar A, Chordia N (2017) Bioinformatics approaches in food sciences. J Food Microbiol Saf Hyg 2:2

Kurilich AC, Tsau GJ, Brown A, Howard L, Klein BP, Jeffery EH et al (1999) Carotene, tocopherol, and ascorbate contents in subspecies of *Brassica oleracea*. J Agric Food Chem 47:1576–1581

Kushad Mosbah M, Brown AF, Kurilich AC, Juvik JA, Klein BP, Wallig MA, Jeffery EH (1999) Variation of glucosinolates in vegetable crops of *Brassica oleracea*. J Agric Food Chem 47: 1541–1548

Kwan CC (1934) Inheritance of some plant characters in cabbage, *Brassica oleracea* var. *capitata*. J Agric Assoc China 126–127:81–127

Lambrix V, Reichelt M, Mitchell-Olds T, Kliebenstein DJ, Gershenzon J (2001) The Arabidopsis epithiospecifier protein promotes the hydrolysis of glucosinolates to nitriles and influences Trichoplusia ni herbivory. Plant Cell 13:2793–2807

Lätari K, Wüst F, Hübner M, Schaub P, Beisel KG, Matsubara S, Beyer P, Welsch R (2015) Tissue-specific apocarotenoid glycosylation contributes to carotenoid homeostasis in *Arabidopsis* leaves. Plant Physiol 168:1550–1562

Lee Y, Lee JY (2019) Blackcurrant (*Ribes nigrum*) extract exerts an anti-inflammatory action by modulating macrophage phenotypes. Nutrients 11. https://doi.org/10.3390/nu11050975

Lee MK, Chun JH, Byeon DH, Chung SO, Park SU, Park S, Arasu MV, Al-Dhabi NA, Lim YP, Kim SJ (2014) Variation of glucosinolates in 62 varieties of Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*) and their antioxidant activity. LWT-Food Sci Technol 58:93–101

Lee YS, Ku KM, Becker TM, Juvik JA (2017) Chemopreventive glucosinolate accumulation in various broccoli and collard tissues: microfluidic-based targeted transcriptomics for by-product valorization. PLoS One 12:e0185112

Li L, Garvin DF (2003) Molecular mapping of *Or*, a gene inducing b-carotene accumulation in cauliflower (*Brassica oleracea* var. *botrytis*). Genome 47:588–594

Li G, Quiros CF (2002) Genetic analysis, expression, and molecular characterization of BoGSL-ELONG, a major gene involved in the aliphatic glucosinolate pathway of *Brassica* species. Genetics 162:1937–1943

Li G, Riaz A, Goyal S, Abel S, Quiros CF (2001) Inheritance of three major genes involved in the synthesis of aliphatic glucosinolates in *Brassica oleracea*. J Am Soc Hortic Sci 126:427–431

Li L, Lu S, O'Halloran DM, Garvin DF, Vrebalov J (2003) High-resolution genetic and physical mapping of the cauliflower high-beta-carotene gene *Or* (Orange). Mol Genet Genomics 270: 132–138

Li L, Lu S, Cosman KM, Earle ED, Garvin DF, O'Neill J (2006) Beta-Carotene accumulation induced by the cauliflower *Or* gene is not due to an increased capacity of biosynthesis. Phytochemistry 67:1177–1184

Li A, Jiang J, Zhang Y, Snowdon RJ, Liang G, Wang Y (2012a) Molecular and cytological characterization of introgression lines in yellow seed derived from somatic hybrids between *Brassica napus* and *Sinapsis alba*. Mol Breed 29:209–219

Li X, Chen L, Hong M, Zhang Y, Zu F, Wen J, Yi B, Ma C, Shen J, Tu J, Fu T (2012b) A large insertion in bHLH transcription factor BrTT8 resulting in yellow seed coat in *Brassica rapa*. PLoS One 7:e44145

Li R, Jiang J, Jia S, Zhu X, Su H, Li J (2019a) Overexpressing broccoli tryptophan biosynthetic genes *BoTSB1* and *BoTSB2* promotes biosynthesis of IAA and indole glucosinolates. Physiol Plant 168:174–187

Li Z, Liu Y, Li L, Fang Z, Yang L, Zhuang M, Zhang Y, Lv H (2019b) Transcriptome reveals the gene expression patterns of sulforaphane metabolism in broccoli florets. PLoS One 14:e0213902

Li Z, Zheng S, Liu Y, Fang Z, Yang L, Zhuang M et al (2021a) Characterization of glucosinolates in 80 broccoli genotypes and different organs using UHPLC-Triple-TOF-MS method. Food Chem 334:127519

Li Z, Liu Y, Yuan S, Han F, Fang Z, Yang L, Zhuang M, Zhang Y, Lv H, Wang Y et al (2021b) Fine mapping of the major QTLs for biochemical variation of sulforaphane in broccoli florets using a DH population. Sci Rep 11:9004

Liu M, Nauta A, Francke C, Siezen RJ (2008) Comparative genomics of enzymes in flavor-forming pathways from amino acids in lactic acid bacteria. Appl Environ Microbiol 74:4590–4600

Liu Z, Hirani A, McVetty PBE, Daayf F, Quiros C, Li G (2012) Reducing progoitrin and enriching glucoraphanin in *Brassica napus* seeds through silencing of the GSL-ALK gene family. Plant Mol Biol 79:179–189

Liu J, Wang WH, Zhang DS, Yu SC, Zhang FL, Zhao XY, Yu YJ, Xu JB, Lu GX (2013) Primary mapping of *pur*, a gene controlling purple leaf color in *Brassica rapa*. Acta Agric Bor-Sin 28: 49–53

Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IAP et al (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun 5:3930

Liu C, Yao X, Li G, Huang L, Wu X, Xie Z (2021) Identification of major loci and candidate genes for anthocyanin biosynthesis in Broccoli using QTL-Seq. Horticulturae 7:246

Liu Y, Du Z, Lin S, Li H, Lu S, Guo L, Tang S (2022) CRISPR/Cas9-targeted mutagenesis of *BnaFAE1* genes confers low-erucic acid in *Brassica napus*. Front Plant Sci 13:848723

Lohr M (2009) Carotenoids. In: Harris EH, Stern DB, Witman GB (eds) The Chlamydomonas sourcebook, 2nd edn. Academic Press, pp 799–817

London SJ, Yuan JM, Chung FL, Gao YT, Coetzee GA, Ross RK, Mimi CY (2000) Isothiocyanates, glutathione S-transferase M1 and T1 polymorphisms, and lung-cancer risk: a prospective study of men in Shanghai, China. Lancet 356:724–729

Longvah T, Anantan I, Bhaskarachary K, Venkaiah K, Longvah T (2017) Indian food composition tables. National Institute of Nutrition, Indian Council of Medical Research, Hyderabad, pp 2–58

Lopez AB et al (2008) Effect of the cauliflower *Or* transgene on carotenoid accumulation and chromoplast formation in transgenic potato tubers. J Exp Bot 59:213–223

Lou P, Zhao J, He H, Hanhart C, Del Carpio DP, Verkerk R, Custers J, Koornneef M, Bonnema G (2008) Quantitative trait loci for glucosinolate accumulation in *Brassica rapa* leaves. New Phytol 179:1017–1032

Lu S, Van Eck J, Zhou X, Lopez AB, O'Halloran DM, Cosman KM et al (2006) The cauliflower *Or* Gene encodes a DnaJ Cysteine-rich domain-containing protein that mediates high-levels of β-Carotene accumulation. Plant Cell 18:3594–3605

Lv H, Wang Y, Han F, Ji J, Fang Z, Zhuang M et al (2020) A high-quality reference genome for cabbage obtained with SMRT reveals novel genomic features and evolutionary characteristics. Sci Rep 10:12394

Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. Nat Rev Genet 10(8):565–577

Macleod G, Macleod AJ (1990) The glucosinolates and aroma volatiles of green kohlrabi. Phytochemistry 29:1183–1187

Magrath R, Bano F, Morgner M, Parkin I, Sharpe A, Lister C, Dean C, Turner J, Lydiate D, Mithen RF (1994) Genetics of aliphatic glucosinolates. I. Side chain elongation in *Brassica napus* and *Arabidopsis thaliana*. Heredity 72:290–299

Magruder R, Myers CH (1933) The inheritance of some plant colours in cabbage. J Agric Res 47: 233–248

Mailer RJ, McFadden A, Ayton J, Redden B (2008) Anti-nutritional components, fibre, sinapine and glucosinolate content, in Australian canola (*Brassica napus* L.) meal. J Am Oil Chem Soc 85:937–944

Mari A, Scala E, Palazzo P, Ridolfi S, Zennaro D et al (2006) Bioinformatics applied to allergy: allergen databases, from collecting sequence information to data integration. The Allergome platform as a model. Cellr Immun 244:97–100

Marwede V, Gul MK, Becker HC, Ecke W (2005) Mapping of QTL controlling tocopherol content in winter oilseed rape. Plant Breed 124:20–26

Matusheski NV, Juvik JA, Jeffery EH (2004) Heating decreases epithiospecifier protein activity and increases sulforaphane formation in broccoli. Phytochemistry 65:1273–1281

Mazumder JA, Khan E, Perwez M et al (2020) Exposure of biosynthesized nanoscale ZnO to *Brassica juncea* crop plant: morphological, biochemical and molecular aspects. Sci Rep 10:8531

McDougall GJ, Fyffe S, Dobson P, Stewart D (2007) Anthocyanins from red cabbage – stability to simulated gastrointestinal digestion. Phytochemistry 68:1285–1294

McNaughton SA, Marks GC (2003) Development of a food composition database for the estimation of dietary intakes of glucosinolates, the biologically active constituents of cruciferous vegetables. Br J Nutr 90:687–697

Meenu AR, Majee M, Pradhan AK, Bisht NC (2015) Genomic origin, expression differentiation and regulation of multiple genes encoding CYP83A1, a key enzyme for core glucosinolate biosynthesis, from the allotetraploid *Brassica juncea*. Planta 241:651–665

Mithen R (2003) Development of isothiocyanate-enriched broccoli, and its enhanced ability to induce phase 2 detoxification enzymes in mammalian cells. Theor Appl Genet 106:727–734

Mithen RF, Clarke J, Lister C, Dean C (1995) Genetics of aliphatic glucosinolates. III. Side chain structure of aliphatic glucosinolates in *Arabidopsis thaliana*. Heredity 74:210–215

Mithen RF, Dekker M, Verkerk R, Rabot S, Johnson IT (2000) The nutritional significance, biosynthesis, and bioavailability of glucosinolates in human foods. J Sci Food Agric 80:967–984

Murovec J, Guc̆ek K, Bohanec B, Avbelj M, Jerala R (2018) DNA-free genome editing of *Brassica oleracea* and *B. rapa* protoplasts using CRISPR-Cas9 ribonucleoprotein complexes. Front Plant Sci 9:1594

Nagarajan J, Ramanan RN, Raghunandan ME, Galanakis CM, Krishnamurthy NP (2017) Carotenoids. In: Galanakis CM (ed) Nutraceutical and functional food components. Academic Press, pp 259–296

Naur P, Petersen BL, Mikkelsen MD, Bak S, Rasmussen H, Olsen CE, Halkier BA (2003) CYP83A1 and CYP83B1, two nonredundant cytochrome P450 enzymes metabolizing oximes in the biosynthesis of glucosinolates in *Arabidopsis*. Plant Physiol 133:63–72

Nesi N, Debeaujon I, Jond C, Pelletier G, Caboche M, Lepiniec L (2000) The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and BAN genes in *Arabidopsis* siliques. Plant Cell 12:1863–1878

Obermuller-Jevic UC, Francz PI, Frank J, Flaccus A, Biesalski HK (1999) Enhancement of UVA induction of haem oxygenase-1 expression by beta-carotene in human skin fibroblasts. FEBS Lett 460:212–216

Okuzaki A, Ogawa T, Koizuka C, Kaneko K, Inaba M, Imamura J, Koizuka N (2018) CRISPR/Cas9-mediated genome editing of the fatty acid desaturase 2 gene in *Brassica napus*. Plant Physiol Biochem 131:63–69

Olinares PDB, Kim J, van Wijk KJ (2011) The Clp protease system; a central component of the chloroplast protease network. Biochim Biophys Acta Bioenerg 1807:999–1011

Omenn GS, Goodman GE, Thornquist MD et al (1996) Effects of a combination of beta-carotene and vitamin A on lung cancer and cardiovascular disease. N Engl J Med 334:1150–1155

Padilla G, Cartea ME, Velasco P, de Haro A, Ordas A (2007) Variation of glucosinolates in vegetable crops of *Brassica rapa*. Phytochemistry 68:536–545

Padmaja LK, Agarwal P, Gupta V, Mukhopadhyay A, Sodhi YS, Pental D, Pradhan AK (2014) Natural mutations in two homoeologous TT8 genes control yellow seed coat trait in allotetraploid *Brassica juncea* (AABB). Theor Appl Genet 127:339–347

Paolini M, Antelli A, Pozzetti L et al (2001) Induction of cytochrome P450 enzymes and over-generation of oxygen radicals in beta-carotene supplemented rats. Carcinogenesis 22:1483–1495

Parkin I, Magrath R, Keith D, Sharpe A, Mithen RF, Lydiate D (1994) Genetics of aliphatic glucosinolates. II. Hydroxylation of alkenyl glucosinolates in *Brassica napus*. Heredity 72:594–598

Parkin IAP, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE et al (2014) Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. Genome Biol 15:r77

Pfalz M, Vogel H, Kroymann J (2009) The gene controlling the indole glucosinolate modifier1 quantitative trait locus alters indole glucosinolate structures and aphid resistance in *Arabidopsis*. Plant Cell 21:985–999

Pfalz M, Mikkelsen MD, Bednarek P, Olsen CE, Halkier BA, Kroymann J (2011) Metabolic engineering in nicotiana benthamiana reveals key enzyme functions in arabidopsis indole glucosinolate modification. Plant Cell 23:716–729

Piccaglia R, Mauro M, Guido B (2002) Factors influencing anthocyanin content in red cabbage (*Brassica oleracea* var. *capitata* L f *rubra* (L) Thell). J Sci Food Agric 82(13):1504–1509

Prasad R, Kumar V, Prasad KS (2014) Nanotechnology in sustainable agriculture: present concerns and future aspects. Afr J Biotechnol 13:705–713

Qiu D, Morgan C, Shi J, Long Y, Liu J, Li R et al (2006) A comparative linkage map of oilseed rape and its use for QTL analysis of seed oil and erucic acid content. Theor Appl Genet 114:67–80

Qu C, Jia L, Fu F, Zhao H, Lu K, Wei L, Xu X, Liang Y, Li S, Wang R, Li J (2017) Genome-wide association mapping and Identification of candidate genes for fatty acid composition in *Brassica napus* L. using SNP markers. BMC Genomics 18(1):232

Rahim A, Afrin KS, Jung HJ, Kim HT, Park JI, Hur Y, Nou IS (2019) Molecular analysis of anthocyanin biosynthesis-related genes reveal BoTT8 associated with purple hypocotyl of broccoli (*Brassica oleracea* var. *italica* L.). Genome 62:253–266

Rao AV, Rao LG (2007) Carotenoids and human health. Pharmacol Res 55:207–216

Reshma V, Adarsh M, Dhanush K (2018) Development of hybrids in cole crops: a review. Plant Arch 18:1–11

Ribaya-Mercado JD, Blumberg JB (2004) Lutein and zeaxanthin and their potential roles in disease prevention. J Am Coll Nutr 23:567S–587S

Richman EL, Carroll PR, Chan JM (2012) Vegetable and fruit intake after diagnosis and risk of prostate cancer progression. Int J Cancer 131:201–210

Riso P, Del Bo' C, Vendrame S (2014) Preventive effects of broccoli bioactives: role on oxidative stress and cancer risk. In: Preedy VR (ed) Cancer: oxidative stress and dietary antioxidants. Elsevier Science Publisher, San Diego, pp 115–126

Rollin P, Tatibouët A (2011) Glucosinolates: the synthetic approach. C R Chim 14:194–210

Rosa EAS, Heaney RK, Fenwick GR, Portas CAM (1997) Glucosinolates in crop plants. Hortic Rev 19:99–215

Ruiz-Sola MA et al (2016) Arabidopsis *GERANYLGERANYL DIPHOSPHATE SYNTHASE 11* is a hub isozyme required for the production of most photosynthesis-related isoprenoids. New Phytol 209:252–264

Rungapamestry V, Duncan AJ, Fuller Z, Ratcliffe B (2006) Changes in glucosinolate concentrations, myrosinase activity, and production of metabolites of glucosinolates in cabbage (*Brassica oleracea* var. *capitata*) cooked for different durations. J Agric Food Chem 54:7628–7634

Salvi S, Tuberosa R (2005) To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci 10:297–304

Sarikamis G, Marquez J, MacCormack R, Bennett RN, Roberts J, Mithen R (2006) High glucosinolate broccoli: a delivery system for sulforaphane. Mol Breed 18:219–228

Sashidhar N, Harloff HJ, Potgieter L, Jung C (2020) Gene editing of three BnITPK genes in tetraploid oilseed rape leads to significant reduction of phytic acid in seeds. Plant Biotechnol J 18:2241–2250

Scalzo RL, Genna A, Branca F, Chedin M, Chassaigne H (2008) Anthocyanin composition of cauliflower (*Brassica oleracea* L. var. *botrytis*) and cabbage (*B. oleracea* L. var. *capitata*) and its stability in relation to thermal treatments. Food Chem 107:136–144

Schmid LM, Ohler L, Mohlmann T, Brachmann A, Muino JM, Leister D, Meurer J, Manavski N (2019) PUMPKIN, the sole plastid UMP kinase, associates with group II introns and alters their metabolism. Plant Physiol 179:248–264

Seow A, Yuan JM, Sun CL, Van Den Berg D, Lee HP, Yu MC (2002) Dietary isothiocyanates, glutathione S-transferase polymorphisms and colorectal cancer risk in the Singapore Chinese Health Study. Carcinogenesis 23:2055–2061

Sharma C, Sadrieh L, Priyani A, Ahmed M, Hassan AH, Hussain A (2011) Anti-carcinogenic effects of sulforaphane in association with its apoptosis-inducing and anti-inflammatory properties in human cervical cancer cells. Cancer Epidemiol 35:272–278

Sharma A, Pandey A, Sharma S, Chatterjee I, Mehrotra R, Sehgal A, Sharma JK (2014) Genetic polymorphism of glutathione S-transferase P1 (GSTP1) in Delhi population and comparison with other global populations. Meta Gene 2:134–142

Shumskaya M, Wurtzel ET (2013) The carotenoid biosynthetic pathway: thinking in all dimensions. Plant Sci 208:58–63

Simopoulos AP (2010) Nutrigenetics/nutrigenomics. Annu Rev Public Health 31:53–68

Singh S, Kalia P (2021) Advances in Cauliflower (*Brassica oleracea* var. *botrytis* L.) breeding with emphasis on India. In: Al-Khayri JM, Jain SM, Johnson DV (eds) Advances in plant breeding strategies: vegetable crops. Springer, Cham, pp 247–301

Singh S, Kalia P, Meena RK, Mangal M, Islam S, Saha S, Tomar BS (2020) Genetics and expression analysis of anthocyanin accumulation in curd portion of Sicilian purple to facilitate biofortification of Indian cauliflower. Front Plant Sci 10:1766

Smith TK, Lund EK, Clarke RG, Bennet RN, Johnson IT (2005) Effects of Brussels sprout juice on the cell cycle and adhesion of human colorectal carcinoma cells (HT29) *in vitro*. J Agric Food Chem 53:3895–3901

Sohail AU, Shad S, Ilyas N, Manaf A, Raja NI, Mashwani ZU (2019) *In vitro* germination and biochemical profiling of *Brassica napus* in response to biosynthesised zinc nanoparticles. IET Nanobiotechnol 13:46–51

Sommerburg O, Keunen JE, Bird AC, Van Kuijk FJ (1998) Fruits and vegetables that are sources for lutein and zeaxanthin: the macular pigment in human eyes. Brit J Ophthalmol 82:907–910

Sønderby IE, Geu-Flores F, Halkier BA (2010) Biosynthesis of glucosinolates – gene discovery and beyond. Trends Plant Sci 15:283–290

Sotelo T, Soengas P, Velasco P, Rodríguez VM, Cartea ME (2014) Identification of metabolic QTLs and candidate genes for glucosinolate synthesis in *Brassica oleracea* leaves, seeds and flower buds. PLoS One 9:e91428

Spitz MR, Duphorne CM, Detry MA, Pillow PC, Amos CI, Lei L, Andrade MD, Gu X, Hong WK, Wu X (2000) Dietary intake of isothiocyanates: evidence of a joint effect with glutathione S-transferase polymorphisms in lung cancer risk. Cancer Epidemiol Biomark Prev 9: 1017–1020

Steck SE, Gaudet MM, Eng SM, Britton JA, Teitelbaum SL, Neugut AI, Santella RM, Gammon MD (2007) Cooked meat and risk of breast cancer: lifetime versus recent dietary intake. Epidemiology 18(3):373–382

Stowe KA, Marquis RJ (2011) Cost of defense: correlated responses to divergent selection for foliar glucosinolate content in *Brassica rapa*. Evol Ecol 25:763–775

Strauch RC, Mengist MF, Pan K, Yousef GG, Iorizzo M, Brown AF, Lila MA (2019) Variation in anthocyanin profiles of 27 genotypes of red cabbage over two growing seasons. Food Chem 301: 125289

Strobbe S, Lepeleire JD, Straeten DVD (2018) From in plant function to vitamin-rich food crops: the ACE of biofortification. Front Plant Sci 9:1862

Sun RF, Zhang SJ, Zhang SF, Li F (2006) Research on creation of purple Chinese cabbage germplasm. Acta Hort Sin 33:1032

Sun B, Liu N, Zhao Y, Yan H, Wang Q (2011) Variation of glucosinolates in three edible parts of Chinese kale (*Brassica alboglabra* Bailey) varieties. Food Chem 124:941–947

Sun T, Yuan H, Cao H, Yazdani M, Tadmor Y, Li L (2018) Carotenoid metabolism in plants: the role of plastids. Mol Plant 11:58–74

Sun D, Wang C, Zhang X, Zhang W, Jiang H, Yao X et al (2019) Draft genome sequence of cauliflower (*Brassica oleracea* L. var. *botrytis*) provides new insights into the C genome in *Brassica* species. Hortic Res 6:82

Suthar PC, Purkait P, Uttaravalli K, Sarkar BN, Ameta R, Sikdar M (2018) Glutathione S-transferase M1 and T1 null genotype frequency distribution among four tribal populations of western India. J Genet 97:11–12

Talalay P, Fahey JW (2001) Phytochemicals from cruciferous plants protect against cancer by modulating carcinogen metabolism. J Nutr 131:3027–3033

Tan H, Wang X, Fei Z et al (2020) Genetic mapping of green curd gene *Gr* in cauliflower. Theor Appl Genet 133:353–364

Tao C, Guan M, Zhou B, Peng Z, Xing M, Wang X, Guan C (2021) Progress of CRISPR/Cas9 technology in breeding of *Brassica napus*. Oil Crop Sci 6:53–57

Teutonico RA, Osborn TC (1994) Mapping of RFLP and qualitative trait loci in *Brassica rapa* and comparison to the linkage maps of *B. napus, B. oleracea*, and *Arabidopsis thaliana*. Theor Appl Genet 89:885–894

Thiruvengadam M, Kim SH, Chung IM (2015) Exogenous phytohormones increase the accumulation of health-promoting metabolites and influence the expression patterns of biosynthesis related genes and biological activity in Chinese cabbage (*Brassica rapa* spp. *pekinensis*). Sci Hortic 193:136–146

Tortorella SM, Royce SG, Licciardi PV, Karagiannis TC (2015) Dietary sulforaphane in cancer chemoprevention: the role of epigenetic regulation and HDAC inhibition. Antioxid Redox Signal 22(16):1382–1424

Traka MH (2016) Health benefits of glucosinolates. Adv Bot Res 80:247–279

Traka M, Mithen RF (2009) Glucosinolates, isothiocyanates and human health. Phytochem Rev 8: 269–282

Traka M, Gasper AV, Melchini A, Bacon JR, Needs PW, Frost V, Chantry A, Jones AM, Ortori CA, Barrett DA, Ball RY (2008) Broccoli consumption interacts with GSTM1 to perturb oncogenic signalling pathways in the prostate. PLoS One 3:e2568

Tzuri G et al (2015) A "golden" SNP in *CmOr* governs the fruit flesh color of melon (*Cucumis melo*). Plant J 82:267–279

Ullah N, Khan J, Khan MW, Raza H, Alam M, Ullah H, Hussain I, Jan IU, Rahman Z, Ali F (2017) Genetic variability for biochemical traits among advanced lines of *Brassica*. Pure Appl Biol 6:1–8

USDA Nutrient Database (2008) USDA National Nutrient Database for Standard Reference (SR)

Uzunova M, Ecke W, Weissleder K, Robbelen G (1995) Mapping the genome of rapeseed (*Brassica napus* L).1. Construction of an RFLP linkage map and localization of QTLs for seed glucosinolate content. Theor Appl Genet 90:194–204

van Doorn HE, van der Kruk GC, van Holst GJ, Raaijmakers-Ruijs N, Postma E, Groeneweg B, Jongen WHF (1998) The glucosinolates sinigrin and progoitrin are important determinants for taste preference and bitterness of Brussels sprouts. J Sci Food Agric 78:30–38

VanEtten CH, Daxenbichler ME, Williams PH, Kwolek F (1976) Glucosinolates and derived products in cruciferous vegetables. Analysis of the edible part from twenty-two varieties of cabbage. J Agric Food Chem 24:452–455

Velasco P, Cartea ME, Gonza'lez C, Vilar M, Ordas A (2007) Factors affecting the glucosinolate content of kale (*Brassica oleracea acephala* group). J Agric Food Chem 55:955–962

Verkerk R, Schreiner M, Krumbein A, Ciska E, Holst B, Rowland I, de Schrijver R, Hansen M, Gerhäuser C, Mithen R, Dekker M (2009) Glucosinolates in *Brassica* vegetables: the influence of the food supply chain on intake, bioavailability and human health. Mol Nutr Food Res 53:219–265

Volden J, Borge GIA, Bengtsson GB, Hansen M, Thygesen IE, Wicklund T (2008) Effect of thermal treatment on glucosinolates and antioxidant-related parameters in red cabbage (*Brassica oleracea* L. ssp. *capitata* f. *rubra*). Food Chem 109:595–605

Walther B, Sieber R (2011) Bioactive proteins and peptides in foods. Int J Vitam Nutr Res 81:181

Wang LI, Giovannucci EL, Hunter D, Neuberg D, Su L, Christiani DC (2004) Dietary intake of Cruciferous vegetables, Glutathione S-transferase (GST) polymorphisms and lung cancer risk in a Caucasian population. Cancer Causes Control 15:977–985

Wang H, Sun S, Liu B, Cheng F, Sun R, Wang X (2011) Glucosinolate biosynthetic genes in *Brassica rapa*. Gene 487:135–142

Wang J, Gu H, Yu H, Zhao Z, Sheng X, Zhang X (2012a) Genotypic variation of glucosinolates in broccoli (*Brassica oleracea* var. *italica*) florets from China. Food Chem 133:735–741

Wang X, Zhang C, Li L, Fritsche S, Endrigkeit J, Zhang W, Long Y, Jung C, Meng J (2012b) Unraveling the genetic basis of seed tocopherol content and composition in rapeseed (*Brassica napus* L.). PLoS One 7:e50038

Wang J, Yu H, Zhao Z, Sheng X, Shen Y, Gu H (2019) Natural variation of glucosinolates and their breakdown products in broccoli (*Brassica oleracea* var. *italica*) seeds. J Agric Food Chem 67: 12528–12537

Wang Y, Ji J, Fang Z, Yang L, Zhuang M, Zhang Y, Lv H (2022) BoGDB: an integrative genomic database for *Brassica oleracea* L. Front Plant Sci 13:852291

Watkins JL, Pogson BJ (2020) Prospects for carotenoid biofortification targeting retention and catabolism. Trends Plant Sci 25:501–512

Wei S, Xiao X, Wei L, Li L, Li G, Liu F et al (2021) Development and comprehensive HS-SPME/GC–MS analysis optimization, comparison, and evaluation of different cabbage cultivars (*Brassica oleracea* L. var. *capitata* L.) volatile components. Food Chem 340:128166

Welsch CW (1985) Host factors affecting the growth of carcinogen-induced rat mammary carcinomas: a review and tribute to Charles Brenton Huggins. Cancer Res 45:3415–3443

Welsch R, Zhou X, Yuan H, Álvarez D, Sun T, Schlossarek D, Yang Y, Shen G, Zhang H, Rodriguez-Concepcion M, Thannhauser TW (2018) Clp protease and OR directly control the proteostasis of phytoene synthase, the crucial enzyme for carotenoid biosynthesis in *Arabidopsis*. Mol Plant 11:149–162

Welsch R, Zhou X, Koschmieder J, Schlossarek T, Yuan H, Sun T, Li L (2020) Characterization of cauliflower OR mutant variants. Front Plant Sci 10:1716

Wiczkowski W, Topolska J, Honke J (2014) Anthocyanins profile and antioxidant capacity of red cabbages are influenced by genotype and vegetation period. J Funct Foods 7:201–211

Wiesner T, He J, Yelensky R, Esteve-Puig R, Botton T, Yeh I, Lipson D, Otto G, Brennan K, Murali R, Garrido M (2014) Kinase fusions are frequent in Spitz tumours and spitzoid melanomas. Nat Commun 5:1–9

Wittstock U, Halkier BA (2000) Cytochrome P450 CYP79A2 from *Arabidopsis thaliana* L. catalyzes the conversion of L-phenylalanine to phenylacetaldoxime in the biosynthesis of benzylglucosinolate. J Biol Chem 275:14659–14666

Wu X, Beecher GR, Holden JM, Haytowitz DB, Gebhardt SF, Prior RL (2006) Concentrations of anthocyanins in common foods in the United States and estimation of normal consumption. J Agric Food Chem 54:4069–4075

Yan C, An G, Zhu T, Zhang W, Zhang L, Peng L, Chen J, Kuang H (2018) Independent activation of the *BoMYB2* gene leading to purple traits in *Brassica oleracea*. Theor Appl Genet 132:895–906

Yarnell SH (1956) Cytogenetics of the vegetable crops II. Crucifers Bot Rev 22:81–166

Yi GE, Robin AH, Yang K, Park JI, Kang JG, Yang TJ, Nou IS (2015) Identification and expression analysis of glucosinolate biosynthetic genes and estimation of glucosinolate contents in edible organs *of Brassica oleracea* subspecies. Molecules 20:13089–13111

Yu J, Zhao M, Wang X et al (2013a) Bolbase: a comprehensive genomics database for *Brassica oleracea*. BMC Genomics 14:664

Yu X, Wang H, Zhong W, Bai J, Liu P et al (2013b) QTL mapping of leafy heads by genome resequencing in the RIL population of *Brassica rapa*. PLoS One 8:e76059

Yu H, Wang J, Sheng X, Zhao Z, Shen Y, Branca F, Gu H (2019) Construction of a high-density genetic map and identification of loci controlling purple sepal trait of flower head in *Brassica oleracea* L. *italica*. BMC Plant Biol 19:228

Yuan Y, Chiu L-W, Li L (2009) Transcriptional regulation of anthocyanin biosynthesis in red cabbage. Planta 230:1141–1153

Yun-Xiang Z, Myung-Ho L, Beom-Seok P, Seung-Beom H, Doo Hwan K (2008) Metabolic engineering of indole glucosinolates in Chinese cabbage plants by expression of *Arabidopsis* CYP79B2, CYP79B3, and CYP83B1. Mol Cells 25:231–241

Zang YX, Kim JH, Park YD, Kim DH, Hong SB (2008a) Metabolic engineering of aliphatic glucosinolates in Chinese cabbage plants expressing *Arabidopsis* MAM1, CYP79F1, and CYP83A1. BMB Rep 41:472–478

Zang YX, Lim MH, Park BS, Hong SB, Kim DH (2008b) Metabolic engineering of indole glucosinolates in Chinese cabbage plants by expression of *Arabidopsis* CYP79B2, CYP79B3, and CYP83B1. Mol Cells 25:231–241

Zang YX, Kim HU, Kim JA, Lim MH, Jin M, Lee SC, Kwon SJ, Lee SI, Hong JK, Park TH et al (2009) Genome-wide identification of glucosinolate synthesis genes in *Brassica rapa*. FEBS J 276:3559–3574

Zhai Y, Yu K, Cai S, Hu L, Amoo O, Xu L, Yang Y, Ma B, Jiao Y, Zhang C, Khan MH (2020) Targeted mutagenesis of BnTT8 homologs controls yellow seed coat development for effective oil production in *Brassica napus* L. Plant Biotechnol J 18:1153–1168

Zhang Y, Talalay P (1994) Anticarcinogenic activities of organic isothiocyanates: chemistry and mechanisms. Cancer Res 54:1976s–1981s

Zhang Y, Kensler TW, Cho CG, Posner GH, Talalay P (1994) Anticarcinogenic activities of sulforaphane and structurally related synthetic norbornyl isothiocyanates. Proc Natl Acad Sci U S A 91:3147–3150

Zhang MK, Zhang LG, Gong ZH, Hui MX (2008) Screening RAPD markers linked to purple trait of Chinese cabbage and its chromosome location. Acta Bot Boreali-Occident Sin 28:901–906

Zhang DS, Zhang FL, Yu YJ, Xu JB, Yu SC, Wang WH (2011) Genetic analysis of leaf color characteristics with purple genes on Chinese cabbage (Chinese cabbage × purple Pak choi). Chin Veg 1:36–38

Zhang B, Hu Z, Zhang Y, Li Y, Zhou S, Chen G (2012) A putative functional MYB transcription factor induced by low temperature regulates anthocyanin biosynthesis in purple kale (*Brassica oleracea* var. *acephala* f. *tricolor*). Plant Cell Rep 31:281–289

Zhang Y, Huai D, Yang Q, Cheng Y, Ma M, Kliebenstein DJ, Zhou Y (2015) Overexpression of three glucosinolate biosynthesis genes in Brassica napus identifies enhanced resistance to *Sclerotinia sclerotiorum* and *Botrytis cinerea*. PLoS One 10:e0140491

Zhang S, Li P, Qian W, Zhang S, Li F, Zhang H, Wang X, Sun R (2016) Mapping and expression profiling reveal an inserted fragment from purple mustard involved anthocyanin accumulation in Chinese cabbage. Euphytica 212:83–95

Zhao B, Seow A, Lee EJD, Poh WT, The M, Eng P, Wang YT, Tan WC, Yu MC, Lee H-P (2001) Dietary isothiocyanates, glutathione S-transferase -M1, −T1 polymorphisms and lung cancer risk among Chinese Women in Singapore. Cancer Epidemiol Biomark Prev 10:1063–1067

Zhao JY, Becker HC, Zhang DQ, Zhang YF, Ecke W (2005) Oil content in a European × Chinese rapeseed population: QTL with additive and epistatic effects and their genotype-environment interactions. Crop Sci 45:51–59

Zhao J, Paulo MJ, Jamar D, Lou P, van Eeuwijk F, Bonnema G, Vreugdenhil D, Koornneef M (2007) Association mapping of leaf traits, flowering time, and phytate content in *Brassica rapa*. Genome 50:963–973

Zhao Y, Yue Z, Zhong X, Lei J, Tao P, Li B (2020) Distribution of primary and secondary metabolites among the leaf layers of headed cabbage (*Brassica oleracea* var. *capitata*). Food Chem 312:126028

Zhou X, Van Eck J, Li L (2008) Use of the cauliflower *Or* gene for improving crop nutritional quality. Biotechnol Ann Rev 14:171–190

Zhou X, Fei Z, Thannhauser TW, Li L (2011a) Transcriptome analysis of ectopic chloroplast development in green curd cauliflower (*Brassica oleracea* L. var. *botrytis*). BMC Plant Biol 11:169

Zhou X, Sun TH, Wang N, Ling HQ, Lu S, Li L (2011b) The cauliflower orange gene enhances petiole elongation by suppressing expression of eukaryotic release factor 1. New Phytol 190:89–100

Zhou X, Welsch R, Yang Y, Álvarez D, Riediger M, Yuan H, Fish T, Liu J, Thannhauser TW, Li L (2015) *Arabidopsis* OR proteins are the major posttranscriptional regulators of phytoene synthase in controlling carotenoid biosynthesis. Proc Natl Acad Sci U S A 112:3558–3563

Zhu B, Wang Z, Yang J, Zhu Z, Wang H (2012) Isolation and expression of glucosinolate synthesis genes CYP83A1 and CYP83B1 in Pak Choi (*Brassica rapa* L. ssp. *chinensis* var. *communis* (N. Tsen & S.H. Lee) Hanelt). Int J Mol Sci 13:5832–5843

Zou J, Jiang C, Cao Z, Li R, Long Y, Chen S, Meng J (2010a) Association mapping of seed oil content in *Brassica napus* and comparison with quantitative trait loci identified from linkage mapping. Genome 53:908–916

Zou J, Zhu JL, Huang SM, Tian ET, Xiao Y, Fu DH et al (2010b) Broadening the avenue of intersubgenomic heterosis in oilseed *Brassica*. Theor Appl Genet 120:283–290

Zuluaga DL, Graham NS, Klinder A, Kloeke AEEVO, Marcotrigiano AR, Wagstaff C, Verkerk R, Sonnante G, Aarts MGM (2019) Overexpression of the MYB29 transcription factor affects aliphatic glucosinolate synthesis in *Brassica oleracea*. Plant Mol Biol 101:65–79

Health-Enhancing Compounds in Carrots: Genetics, Genomics, and Molecular Breeding

Pablo F. Cavagnaro, Frank Dunemann, Raman Selvakumar, Massimo Iorizzo, and Philipp W. Simon

Contents

P. F. Cavagnaro (✉)
National Scientific and Technical Research Council (CONICET), National Institute of Agricultural Technology (INTA) E.E.A. La Consulta, La Consulta CC8, San Carlos, Mendoza, Argentina
e-mail: cavagnaro.pablo@inta.gob.ar

F. Dunemann
Julius Kühn-Institut (JKI), Federal Research Centre for Cultivated Plants, Institute for Breeding Research on Horticultural Crops, Quedlinburg, Germany

R. Selvakumar
Centre for Protected Cultivation Technology, ICAR-Indian Agricultural Research Institute (IARI), New Delhi, India

M. Iorizzo
Plants for Human Health Institute and Department of Horticultural Sciences, North Carolina State University, Kannapolis, NC, USA

P. W. Simon
USDA Agricultural Research Service, Vegetable Crops Research Unit, Department of Horticulture, University of Wisconsin, Madison, WI, USA

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_51

Abstract

Carrot is an economically important vegetable crop worldwide. Its storage root, the consumed organ, varies broadly within the carrot germplasm, exhibiting different colors due to the accumulation of anthocyanin and carotenoid pigments, as well as extensive variation for phytochemicals composition and consumer-quality traits. Anthocyanins and other phenolics, carotenoids, polyacetylenes, and terpenes represent the major carrot nutraceutical classes. In recent years, the use of next-generation sequencing technologies has facilitated the application of "multi-omics" approaches, in combination with transgenics and classical genetic tools, for studying the genetics underlying the accumulation of these phytochemicals in the carrot root. In purple carrot, such approaches allowed the identification and mapping of simply inherited and quantitative trait loci (QTLs) conditioning anthocyanin pigmentation in different tissues and genetic backgrounds, and the discovery of key genes conditioning anthocyanin biosynthesis, glycosylation, and acylation. Glycosylation and acylation influence the chemical stability and bioavailability of anthocyanins, and therefore their potential use as food colorants or nutraceutical agents, respectively. Similarly, important advances were made for two major loci conditioning carotenoids accumulation in white, yellow, and orange roots, namely *Y* and Y_2. With the sequencing of the carrot genome, a candidate for the *Y* gene involved in photosystem development and carotenoid storage was described, whereas fine mapping of Y_2 drastically reduced the genomic region of interest to 650-kb, but a clear candidate was not identified. Another gene, *Or*, which regulates chromoplasts development, was associated with carotenoids presence in the carrot root. Besides these nonstructural genes, progress towards understanding the role of several carotenoid biosynthetic genes has been made. The genetics of carrot polyacetylenes is also becoming increasingly understood. Candidate fatty acid desaturase 2 (FAD2) genes with specific desaturase and/or acetylenase activities have been identified by QTLs analysis and proposed as catalyzers of different steps in the polyacetylene pathway, and their genomic organization was described. Similarly, gene members of the large

carrot terpene synthase family were catalogued, partially associated with QTLs for characteristic carrot root monoterpenes like sabinene, and functionally characterized in vitro. This chapter reviews and discusses recent advances in genetics and genomics of the main carrot nutraceuticals.

Keywords

Daucus carota · Multi-omics · Anthocyanins · Phenolics · Carotenoids · Polyacetylenes · Terpenes

Abbreviations

AA	Acylated anthocyanins
ABTS	2,2′-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid)
AOC	Antioxidant capacity
Cy3XCGG	Cyanidin-3-(2″-xylose-6″-(4-coumuroyl)glucose-galactoside)
Cy3XFGG	Cyanidin-3-(2″-xylose-6″-feruloyl-glucose-galactoside)
Cy3XG	Cyanidin-3-(2″-xylose-galactoside)
Cy3XGG	Cyanidin-3-(2″-xylose-6-glucose-galactoside)
Cy3XSGG	Cyanidin-3-(2″-xylose-6″-sinapoyl-glucose-galactoside)
DPPH	2,2-diphenyl-1-picrylhydrazyl
HCA	Hydroxycinnamic acid
NAA	Non-acylated anthocyanins
PAC	Provitamin A carotenoid
QTL/QTLs	Quantitative trait locus/loci
TF	Transcription factor
TPC	Total phenolics content

1 General Introduction

The carrot (*Daucus carota* L.) germplasm displays broad genetic and phenotypic variation for many agronomic, compositional, and consumer-quality traits. Its storage root, the main consumed organ, can exhibit different colors due to the accumulation of different combinations of anthocyanin and carotenoid pigments, resulting in anthocyanin-rich purple or black carrots, orange carrots with high levels of β-carotene and α-carotene, both provitamin A carotenoids (PACs), lycopene-rich red carrots, yellow carrots that predominantly accumulate xanthophylls, and white-rooted carrots with nearly undetectable levels of the previous pigments. Each of these pigments has particular physicochemical properties, bioactivities, and health-beneficial effects. Among the latter, strong antioxidant, anti-inflammatory, anticancer, immunomodulatory, cardioprotective, and neuroprotective effects, among others, have been reported for some of these pigments. In addition, carrots of all root colors accumulate varying types and concentrations of other non-anthocyanin phenolics, polyacetylenes (with predominance of the C_{17}-compounds falcarinol, falcarindiol, and falcarindiol-3-acetate), and terpenes. While major carrot phenolics (e.g., chlorogenic acid) are

recognized antioxidant agents, some polyacetylenes, particularly C_{17}-falcarins, have been reported to exert strong anti-inflammatory and anticancer effects against various cancer types. However, in high concentrations, some polyacetylenes can be toxic for human health and may confer unpleasant bitter-off taste to carrots. Terpenoids from carrot essential oils exhibit antimicrobial, antioxidant, and anticancer activities. Together, anthocyanins and other phenolics, carotenoids, polyacetylenes, and terpenes represent the major nutraceutical type classes present in the carrot root.

For each of these carrot phytochemical groups, important advances have been made with regards to characterizing their chemical diversity at various levels, ranging from intra-plant, organ- and tissue-specific compositional variations, to variation across carrot genetic backgrounds in large germplasm collections, to evaluating changes in compounds content and profile under abiotic and biotic stimuli/stressors. In addition, physicochemical, bioactive, and health-related attributes of some of these carrot compounds have been studied and described. Breeding has also been important for some of these traits, mainly associated with pigment content and flavor, resulting – for example – in the development of high PAC carrots with virtually no bitter-off taste, and released cultivars with other root colors.

In recent years, the sequencing of the carrot genome and the improvement in throughputness, cost per sequence read, and availability of next-generation sequencing (NGS) technologies have facilitated the application of "multi-omics" approaches, in combination with transgenics and classical genetic tools, for studying the genetic regulation underlying the accumulation of these phytochemicals in the carrot root. Although progress in understanding the genetic basis of these traits has advanced at varying speeds for the different carrot nutraceuticals, perhaps being most accelerated for anthocyanins and carotenoids, such approaches have allowed significant advances for all the compound types. As a result, in just a few years, it was possible to identify and genetically map simply inherited and QTLs conditioning the type, content, and composition of these phytochemicals across different plant tissues and genetic backgrounds, as well as discovering and functionally characterizing candidate regulatory and structural genes for these traits. This chapter reviewed and discussed recent advances in diversity, health properties, genetics, and genomics of the main carrot nutraceuticals.

2 Anthocyanins

2.1 Introduction

Anthocyanins are water-soluble pigments that confer blue, purple, and red color to several organs and tissues in many plants. In the plant, they serve multiple functions including the attraction of pollinators, and attenuation of biotic and abiotic stresses such as insect and phytopathogen attacks, salinity, drought, low temperatures, exposure to ultraviolet light, heavy metal toxicity, and mechanical wounding, among others (Shirley 1996). Because of their widespread occurrence in many plant foods, predominantly found in higher concentrations in fruits and vegetables, these pigments represent important components of the human diet. Anthocyanin consumption has been linked

to various health benefits, including reduced risk of cardiovascular disease, prevention or delay of cognitive decline and neurodegenerative-related disorders, improved regulation of glycemia, prevention of some autoimmune diseases, and decreased risk of some cancer types (reviewed by He and Giusti 2010). These health benefits are largely attributed to the proven antioxidant and anti-inflammatory properties of these pigments, as it is currently thought that chronic inflammation and oxidative stress represent major components in the etiology of chronic diseases (He and Giusti 2010).

Chemically, anthocyanins belong to the flavonoid subclass of phenolic compounds, and they are produced via the phenylpropanoid pathway, a late branch of the shikimate pathway (Herrmann and Weaver 1999) (Fig. 1). Their structure consists of a central molecule called "anthocyanidin" to which different sugars and organic acids can be attached. Although several anthocyanidins have been described, only six of them are widely found in nature, namely cyanidin, petunidin, pelargonidin, malvidin, delphinidin, and peonidin (Andersen and Jordheim 2006). When anthocyanidins are bound to sugars, the form most commonly found in nature, they are called "anthocyanins." A fairly large list of simple sugars may glycosylate anthocyanidins to form anthocyanins. In addition, the sugars bound to the anthocyanidins are often acylated by different aromatic or aliphatic organic acids. As result, a huge molecular diversity spectrum has been reported in the plant kingdom for these pigments as consequence of the many different combinations found among anthocyanidins, sugars, and organic acids (Andersen and Jordheim 2006). The extent to which these pigments are bound to sugars, a process called "glycosylation," and to organic acids, termed "acylation," strongly influences anthocyanin physicochemical and biological properties (reviewed by Prior and Wu 2006). For instance, glycosylation increases water solubility and chemical stability, whereas acylation of the glycoside residues has been found to confer protection against degradation and significantly increase the chemical stability of these pigments (Mazza et al. 2004). The latter was recently shown in a study by Perez et al. (2022) using carrot anthocyanins, demonstrating the acylated anthocyanins (AA) were significantly less prone to degradation than their non-acylated (NAA) counterparts under higher temperature and pH conditions. Because of their greater chemical stability, AA are preferred over NAA for their potential use as natural food colorants. In contrast, carrot NAA have been shown to be significantly more bioavailable than AA (Kurilich et al. 2005; Charron et al. 2009), and this may be relevant when breeding for increased functional value in anthocyanin-rich crops, as bioavailability is a major factor conditioning the potential of a dietary agent to provide health benefits.

The biosynthetic pathways and genetic regulation of anthocyanins have been studied in extensive detail in several model species – e.g., Arabidopsis, snapdragon (*Antirrhinum majus*), petunia (*Petunia hybrida*), and maize (*Zea mays*) – generally revealing that both structural and regulatory genes participate, yet finding broad variations regarding the key genes that control this process among species (Holton and Cornish 1995). In the last decades, important advances have been made concerning the composition, distribution, physicochemical properties, bioactivities, and – most intensively in recent years – the genetic regulation of anthocyanin pigmentation in carrot. Thus, the following sections concern these aspects of carrot anthocyanin pigmentation and their relevance for breeding and genetic research.

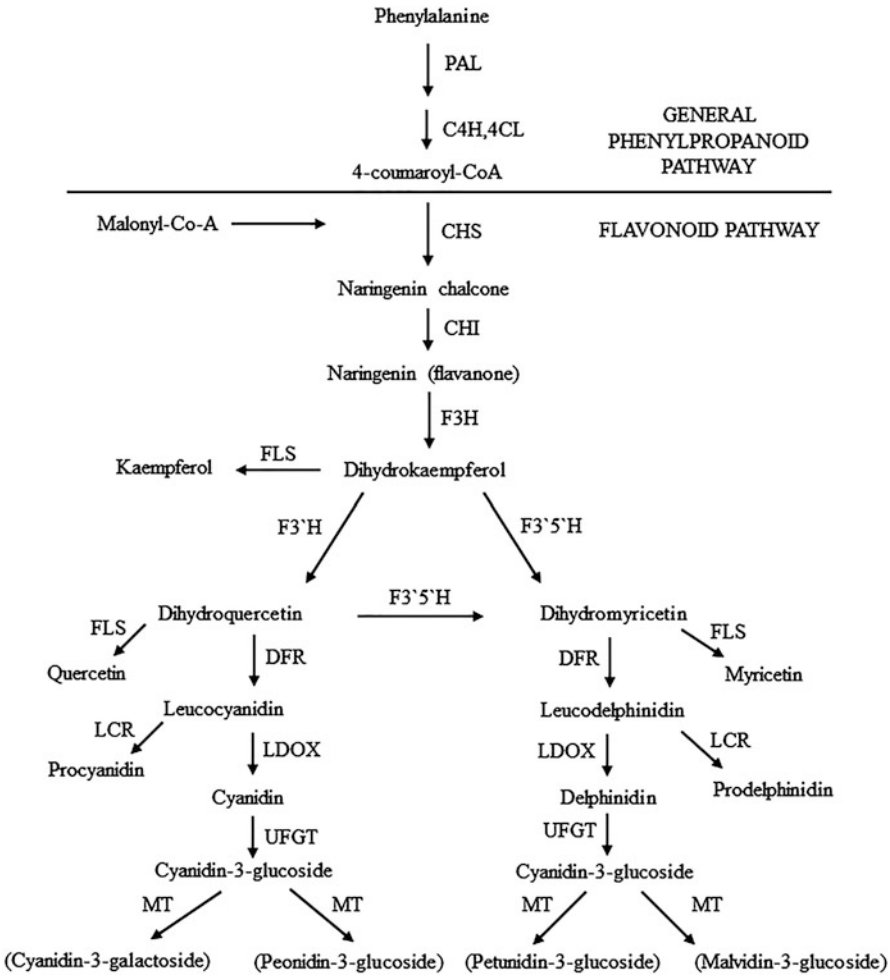


Fig. 1 Schematic representation of the anthocyanin biosynthesis pathway. *PAL* phenylalanine ammonia-lyase, *C4H* cinnamate 4-hydroxylase, *4CL* 4-coumarate CoA ligase, *CHS* chalcone synthase, *CHI* chalcone isomerase, *F3H* flavanone 3-hydroxylase, *FLS* flavonol synthase, *F3′H* flavonoid 3′- hydroxylase, *F3′5′H* flavonoid 3′,5′-hydroxylase, *DFR* dihydroflavonol 4-reductase, *LDOX* leucoanthocyanidin dioxygenase, *UFGT* UDP-glucose-flavonoid 3-*O*-glucosyl-transferase, and *MT* methyltransferase (Reproduced from Cavagnaro and Iorizzo (2019))

2.2 Chemical Diversity of Carrot Anthocyanins and Relevance as Nutraceuticals and Food Colorants

2.2.1 Diversity for Anthocyanin Content and Composition in the Purple Carrot Germplasm

Extensive genetic variation exists among purple carrot genetic stocks for anthocyanin pigmentation across different tissues of the taproot, and tissues and organs of

the aerial plant part (Cavagnaro et al. 2019). Root pigmentation phenotypes across root tissues vary broadly, from dark purple or black fully pigmented roots (i.e., all tissues accumulate anthocyanins in large quantities) to anthocyanin accumulation restricted to the outermost epidermal cell layers, and include intermediate phenotypes exhibiting different intensities and patterns of purple pigmentation in the epidermis, outer phloem (or "cortex"), inner phloem, and xylem tissues. In some particular backgrounds, the color of the root surface is not uniform purple, and anthocyanin pigmentation in this phenotype was estimated visually as a percentage of the root surface (Cavagnaro et al. 2014). As result of these different pigmentation patterns, root total anthocyanin concentration varies broadly across genetic stocks. A first study by Kammerer et al. (2004) reported a range of 45–17,400 mg/kg dry weight (dw) in a collection of 15 purple carrot accessions [in fresh weight (fw), these estimates are equivalent to ~0.5–191 mg/100 g fw, considering 11% of root dry matter]. Later, Montilla et al. (2011) found a range of 1.5–17.7 mg/100 g fresh weight (fw) in 4 commercial cultivars, whereas Algarra et al. (2014) found a range of 93.4–126.4 mg/100 g fw in 2 Spanish cultivars, and Bannoud et al. (2018) 1–229 mg/100 g fw in 26 accessions from diverse geographical origins. Interestingly, there is a strong positive association between the intensity of the purple color (estimated visually) across root tissues and the total anthocyanin concentration, with accessions with dark fully pigmented roots generally presenting the greatest concentration and accessions with pale purple color in the outer phloem and periderm tissues typically exhibiting very low pigment levels. Such association between pigment levels and visual color intensity may be useful for rapidly and inexpensively estimating pigment levels across genetic stocks in purple carrot breeding programs. Also, total anthocyanin content is significantly and positively correlated with TPC, with correlation values (r) ranging from 0.85 (Bannoud et al. 2018) to 0.99 (Leja et al. 2013), suggesting that anthocyanins account for a large proportion of the phenolic compounds present in purple carrots.

The main carrot anthocyanins are cyanidin derivatives (Kammerer et al. 2004; Montilla et al. 2011; Algarra et al. 2014; Bannoud et al. 2018), although pelargonidin, petunidin, and peonidin have been found in trace amounts in some genetic backgrounds (Kammerer et al. 2004; Montilla et al. 2011; Algarra et al. 2014). Among the cyaniding glycosides, two NAA and three AA are commonly found in purple carrots (Table 1). The percentage of AA relative to the total

Table 1 Carrot cyanidin glycosides with approximate HPLC retention times and molecular masses

Compound	Abbreviation	RT	MW
Cy-3-(2″-xylose-6-glucose-galactoside)	Cy3XGG	14.0	743
Cy-3-(2″-xylose-galactoside)	Cy3XG	15.1	581
Cy-3-(2″-xylose-6″-sinapoyl-glucose-galactoside)	Cy3XSGG	15.4	949
Cy-3-(2″-xylose-6″-feruloyl-glucose-galactoside	Cy3XFGG	16.0	919
Cy-3-(2″-xylose-6″-(4-coumuroyl)glucose-galactoside)	Cy3XCGG	16.4	889

RT is retention time (min) for the chromatographic procedure described by Kurilich et al. (2005). MW is molecular weight

anthocyanin content found across accessions in different studies ranged from 25% to 99% (Kammerer et al. 2004; Netzel et al. 2007; Montilla et al. 2011; Algarra et al. 2014, Cavagnaro et al. 2014), although in most commercial cultivars AA generally predominate over NAA. In absolute values, carrot roots may have up to ~155 mg/100 g fw of AA (Kammerer et al. 2004).

Purple carrots vary in anthocyanin composition and relative content (%) of individual pigments. In general, cyanidin glycosides acylated with ferulic acid (Cy3XFGG) predominate in most accessions, followed by glycosides of the same anthocyanidin acylated with sinapic (Cy3XSGG) and coumaric acid (Cy3XCGG). Evaluation of anthocyanin composition in 15 purple carrots revealed that 13 accessions had Cy3XFGG as the most abundant pigment, representing its concentration 42.5–83.8% of the total anthocyanin content, whereas Cy3XSGG predominated in the other 2 accessions (Kammerer et al. 2004). Another study that examined four purple-rooted cultivars found Cy3XFGG as the major pigment in three accessions and Cy3XSGG in the remaining one (Montilla et al. 2011). Coincidently, Bannoud et al. (2019) characterized pigment composition in a carrot F_2 mapping population segregating for purple pigmentation in different root tissues, reporting Cy3XFGG and Cy3XSGG (in that order) as the predominant anthocyanins in both the phloem and xylem. Non-acylated glycosides were in general less abundant, with Cy3XG usually predominating over Cy3XGG (Kammerer et al. 2004; Montilla et al. 2011, Bannoud et al. 2019).

2.2.2 Acylated Anthocyanins for the Food Colorant Industry and Non-acylated Anthocyanins for Increased Nutritional Functionality

A major disadvantage of natural colorants relative to synthetic dyes is, generally, that the formers are more susceptible to degradation. In contrast, synthetic dyes tend to be chemically more stable and maintain their color across a broad range of food conservation conditions, tolerating greater variations in, for example, temperature, acidity, oxygen concentration, and light intensity. Despite their generally higher chemical stability, in the last decades, a number of potential health risks have been associated with the consumption of synthetic food dyes, including behavioral and neurological disorders (McCann et al. 2007), higher incidence of allergies (Koutsogeorgopoulou et al. 1998), and potential carcinogenesis (Dees et al. 1997). Such health-related concerns have raised the interest in finding alternative sources of natural colorants for the food industry.

A very recent study by Perez et al. (2022) investigated the chemical stability of anthocyanins from purple carrots as compared to that of a synthetic food dye (E131) and a commercial food colorant from grape, under a range of temperature and pH conditions. It was found that carrot anthocyanin extracts had greater thermal stability – with regards to color intensity, color space parameters, and concentration of anthocyanins – than anthocyanins from the grape extract and the synthetic dye. The higher stability of the carrot colorants was attributed to their richness in AA. Also, carrot anthocyanins, in general, were more stable at low pH and temperature conditions (i.e., 4 °C, pH = 2.5), whereas under less favorable conditions

(25–40 °C, pH = 7.0), the rate of decay was substantially and significantly different among the individual anthocyanins, with all AA being less affected than the NAA. In particular, anthocyanins acylated with feruloyl (Cy3XFGG) and coumaroyl glycosides (Cy3XCGG) were the most stable carrot pigments. Altogether, these data strongly suggest that carrots with high levels of AA and, in particular, Cy3XFGG and Cy3XCGG, would be most suitable as sources of colorants for the food industry.

Based on these previous data, and from the perspective of developing suitable plant materials for the production of food colorants, three objectives appear as highly relevant: (i) increasing total anthocyanin concentration in the root; (ii) increasing the AA:NAA ratio as much as possible, ideally to the extent that AA account for nearly all of the root anthocyanins; and (iii) shift the AA composition towards a predominance of Cy3XFGGand Cy3XCGG (the most chemically stable AAs) over Cy3XSGG (the least stable AA). A comprehensive review on the potential of carrot anthocyanins as food colorants was recently published by Iorizzo et al. (2020). From another point of view, carrots rich in NAA would be ideal for increasing anthocyanin bioavailability (Kurilich et al. 2005; Charron et al. 2009) and – therefore – nutritional value in, for instance, fresh-consumption carrots, although genotypes with predominance of NAA are not frequent in the carrot germplasm. Considering these two breeding goal (i.e., food colorant production and fresh-consumption carrots with high pigment bioavailability), it is advisable to devote efforts towards characterizing anthocyanin profiles in larger purple carrot germplasm collections, to identify suitable genotypes for either purpose, and to advance research on the genetic regulation of these pigments.

2.2.3 Nutraceutical Properties of Carrot Anthocyanins

Several studies have addressed nutritional aspects of purple carrot anthocyanins. In a clinical feeding trial conducted by Kurilich et al. (2005), fresh and cooked carrots rich in anthocyanins were consumed by human subjects, and the recovery rate of anthocyanin pigments in the blood and urine was estimated. It was found that the recovery of NAA was 8–14 times higher than recovery of AA, whereas the consumption of cooked carrots increased further the recovery rate of NAA but not their acylated counterparts. In this study, whole carrots were use and, therefore, it was not possible to separate the effect of the anthocyanin molecular structure from that of the root matrix on pigment bioavailability. Thus, in a following study by Charron et al. (2009) purple carrot juice, instead of whole carrots, was used in order circumvent matrix effects, revealing recovery rates of NAA nearly fourfold higher than those of AA. Given that anthocyanins were consumed as a carrot juice, without the root matrix, the higher bioavailability found in NAA is mainly attributable to their molecular structure, rather than to possible interactions with the food matrix. In line with these results, Netzel et al. (2007) reported that after consumption of an anthocyanin rich carrot extract by human subjects, the urine recovery of NAA was eightfold higher than that of AA.

Sun et al. (2009) analyzed carrot antioxidant capacity (AOC), using ABTS and DPPH-based procedures, in seven carrot cultivars with different root color, finding that purple carrots had the greatest AOC, with values 3.6- to 28-fold higher than the

AOC in other root colors. The high AOC of purple-rooted carrots was associated with their high concentration of anthocyanins and other phenolics but not with carotenoids levels. The latter contributed minimally to the overall AOC of purple carrots, representing less than 3% of the total AOC. In agreement with results from Sun et al., Leja et al. (2013) reported that purple carrots (two genotypes) were the plant materials with greatest AOC in a collection of 35 carrot cultivars with different root color, noting also that the average TPC content in purple-rooted carrots was ~ninefold greater than in cultivars of other root colors, suggesting that TPC levels may explain the observed variation in AOC. Coincidently, AOC analysis in two purple-rooted and one orange carrot revealed significantly greater AOC in purple carrots, with AOC being directly correlated with root anthocyanin level.

Narayan et al. (1999) reported potent antioxidant activity of purple carrot anthocyanins, as indicated by their ability to inhibit enzymatic and nonenzymatic lipid peroxidation in vitro, in a dose-dependent manner. Furthermore, under this system, carrot anthocyanins exhibited stronger AOC than the classical food antioxidant preservatives "butylated hydroxyanisole" (BHA) (food additive E320) and "butylated hydroxytoluene" (BHT). In addition, carrot anthocyanins were found to confer protection against oxidative stress in human colon mucosa cells, reducing oxidative DNA damage in ~20% relative to untreated (control) cells. Interestingly, this study used the fraction of the purple carrot extract that passed through a simulated gastrointestinal tract (including the stomach, small intestine, and colon), suggesting that even after partial degradation by digestion, carrot anthocyanins exerted significant antioxidant effects in the colon. In full agreement with these data, anthocyanin-rich phenolic extracts from purple carrots exhibited strong AOC in a colon carcinoma (Caco-2) cell line, by means of direct radical scavenging and reduction of H_2O_2-induced inflammation markers, such as interleukins (IL)-1β, IL-6, IL-8, and tumor necrosis factor (TNF)-α (Zhang et al. 2016). In addition, the carrot anthocyanin extracts increased the activity of endogenous antioxidant enzymes and glutathione concentration.

In addition to the antioxidant and anti-inflammatory effects, these carrot pigments may confer protection against some cancer types. The use in vitro of carrot extracts rich in anthocyanins on human cancer cell lines (promyelocytic leukemia HL-60 and colorectal adenocarcinoma HT-29) revealed significant and dose-dependent antiproliferative effects in both types of cancer cells (Netzel et al. 2007). Jing et al. (2008) examined the antiproliferative effects of various anthocyanin-rich extracts obtained from different fruits and vegetables (purple carrot, bilberry, purple corn, chokeberry, grape, elderberry, and radish) varying in the chemical structure of their anthocyanins on a cell line of human colorectal adenocarcinoma (HT29), reporting that all anthocyanin extracts significantly inhibited HT29 growth, although with varying antiproliferative potencies: purple corn > chokeberry and bilberry > purple carrot and grape > radish and elderberry. Interestingly, monoglycosylated NAA exerted the strongest antiproliferative effects, whereas triglycosylated AA had the least effect. Another study by Sevimli-Gur et al. (2013) expanded the anticancer *in vitro* evaluation of purple carrot extracts to various cancer cell lines from human (MCF-7, SK-BR-3, and MDA-MB-231 for breast cancer; HT-29 for colon cancer;

and PC-3 for prostate cancer) and mouse (Neuro 2A for neuroblastoma), reporting potent and dose-dependent cytotoxic effects on all of these cancer lines, with the strongest antiproliferative effects found in brain neuroblastoma. Considering that the study also found very little cytotoxicity by the carrot extract in a normal – not cancerous – cell line (VERO, from African green monkey kidney), the authors concluded that purple carrot extracts may have greatest potential as a dietary treatment for brain cancer, avoiding negative side effects in normal cells.

2.3 Genetics and Genes Controlling Carrot Anthocyanin Pigmentation

2.3.1 Inheritance and Mapping of Simply Inherited Traits

A first inheritance study by Simon (1996) described a major effect locus, termed P_1, controlling the presence or absence of purple color in the taproot of carrot, being purple dominant over non-purple, by analyzing segregation ratios in different populations derived from crosses between purple and non-purple rooted carrots. Purple pigmentation was also scored in aerial plant parts of the same segregating populations, leading to the discovery of another simply inherited dominant locus, P_2, which conditioned pigmentation in the nodes, and it was estimated that both loci were genetically linked at a distance of ~36 cM. In two subsequent studies by Vivek and Simon (1999) and Yildiz et al. (2013), P_1 was mapped to carrot chromosome 3 in F_2 mapping populations derived from crosses that used "B7262" as the purple root progenitor. B7262 is a Turkish carrot with anthocyanin accumulation only in the outer phloem of the root, being the inner phloem and xylem orange. A later study by Cavagnaro et al. (2014) identified and mapped another simply inherited dominant locus, called P_3, controlling anthocyanin accumulation in the taproot and in leaf petioles in the genetic backgrounds of P9547 (from Turkey) and PI652188 (from China), both exhibiting purple color in the taproot and petioles. P_3 mapped to Chr. 3 but it was positionally unrelated to P_1, distanced from the latter at >30 cM, as indicated from results of comparative map analysis using common markers across the three segregating populations were the two traits had been mapped (Cavagnaro et al. 2014). Figure 2b presents root phenotypes of the purple-rooted progenitors and derivative segregating populations used to map these and other simply inherited and QTLs conditioning root and petiole anthocyanin pigmentation.

More recently, Iorizzo et al. (2019) analyzed segregation for anthocyanin pigmentation in the root and petioles in an F_2 mapping population derived from BP85682, a Syrian carrot with purple root and petioles, and in advanced progenies (F_3 and F_5 families) of the populations used earlier by Cavagnaro et al. (2014). Phenotypic segregations for root and petiole anthocyanin pigmentation co-segregated in the F_2, and both traits exhibited a 3:1 ratio (purple:non-purple), suggesting that a single major-effect locus conditions both traits. Comparative linkage mapping using markers in common across this population and other populations with known pedigree revealed that this anthocyanin-conditioning locus in the Syrian BP85682 background corresponds to P_3. Thus, the phenotypic

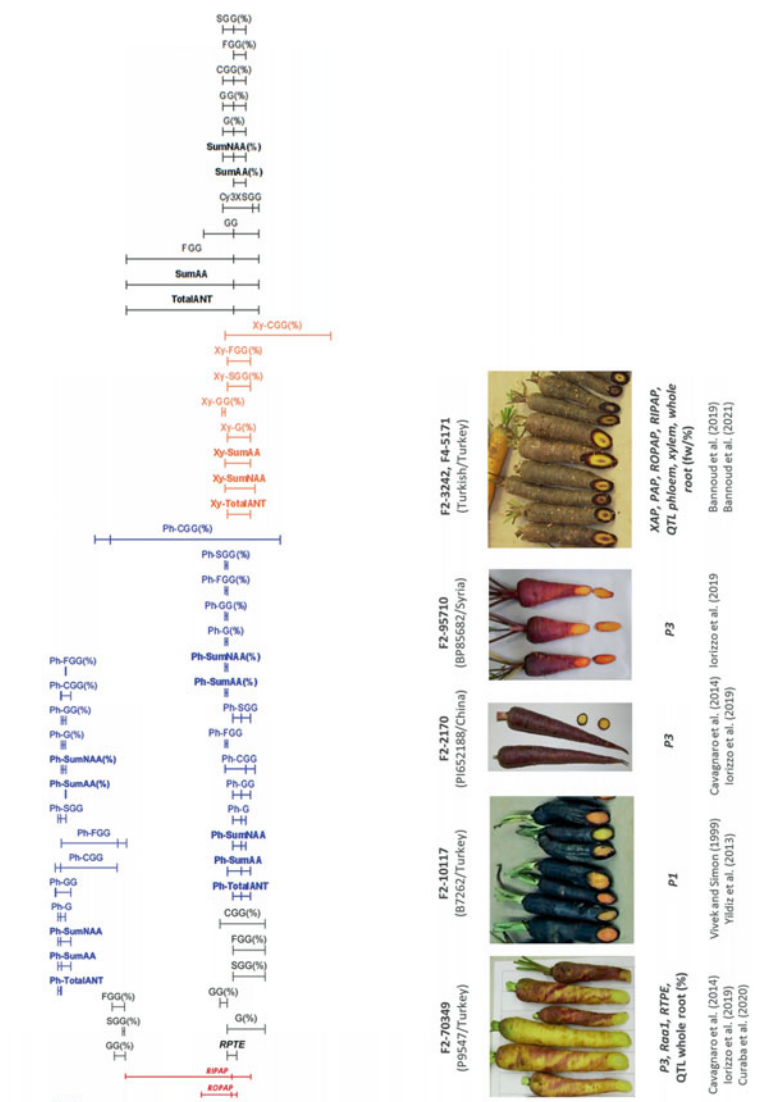


Fig. 2 Genomic locations of anthocyanin QTLs in chromosome 3 and phenotypes mapped in carrot with respective references. (**a**) Representative QTLs and simply inherited traits conditioning the presence and concentration of anthocyanins mapped onto carrot chromosome 3, displaying the physical position of genes associated with anthocyanin biosynthesis. Transcription factors are indicated in purple; anthocyanin biosynthetic structural genes in black; anthocyanin acyltransferases, glucosyltransferases, and methyltransferases in green; and genes involved in anthocyanins cellular transport in orange. The genes were labeled

locus P_3 conditions root and petiole anthocyanin pigmentation in all of the purple carrot materials evaluated to date, except for the Turkish carrot B7262, in which P_1 controls root but not petiole pigmentation.

Bannoud et al. (2019) investigated tissue-specific anthocyanin pigmentation in the carrot root and leaves, using F_2 and F_3 populations segregating for purple color in the root phloem and xylem tissues, and in the leaf petiole. They described and mapped two simply inherited loci controlling the presence/absence of pigmentation in the root xylem and leaf petioles, with purple color being dominant over non-purple in both tissues. These loci, called *XAP* and *PAP*, for "xylem anthocyanin pigmentation" and "petiole anthocyanin pigmentation," respectively, were mapped in the same chromosome region of P_3 (Fig. 2a). In this genetic background, purple phloem pigmentation segregated consistently with a two-gene model, dominant for purple over non-purple, with its main effects co-localizing with the P_1 and P_3 regions. In a recent follow-up study concerning tissue-specific pigmentation in the root outer phloem (also called "cortex") and inner phloem tissues, Bannoud et al. (2021) genetically mapped purple pigmentation in both the outer and inner phloem tissues. The mapping of these simply inherited phenotypic traits, called *ROPAP* and *RIPAP*, for "root outer-phloem and root inner-phloem anthocyanin pigmentation," respectively, revealed colocalization of *ROPAP* with the P_1 and P_3 regions, whereas *RIPAP* co-localized with P_3 only (Fig. 2a). Noteworthy, the populations used by Bannoud et al. had the unprecedented characteristic that P_1 and P_3 as well as pigmentation in the root phloem and xylem, and in leaf petioles all segregated in the same mapping population (3242). Together, the studies of Bannoud et al. (2019, 2021) provide information on the tissue-specificity of P_1 and P_3, and reinforce previous findings by Cavagnaro et al. (2014) and Iorizzo et al. (2019) suggesting that, in some genetic backgrounds, P_3 controls root and petiole anthocyanin pigmentation, while in other backgrounds (e.g., B7262), P_1 controls pigmentation exclusively in the root outer phloem (Yildiz et al. 2013).

Fig. 2 (continued) by their abbreviated names followed by the DCAR or LOC number, in parenthesis. The physical position of each gene in the chromosomes is expressed in terms of nucleotide coordinates from the carrot genome assembly and indicated by the ruler on the left. Major-effect phenotypic traits are denoted in red, italic, and bold fonts. QTLs conditioning absolute (i.e., expressed on a fresh weight basis) or relative pigment concentration (i.e., % of the total anthocyanin content) in the whole root (in black), as well as in the root phloem (in blue) or xylem tissues (in orange), are presented. QTLs for total or combined anthocyanin pigments (e.g., sum of acylated anthocyanins) are indicated in bold. QTLs bars indicate the 1.5 LOD interval (nt) and the position of the maximum LOD value. QTLs are labelled by their pigment abbreviations G = Cy3XG; GG = Cy3XGG; CGG = Cy3XCGG; FGG = Cy3XFGG; SGG = Cy3XSGG; Total ANT = total anthocyanins; Sum AA = sum of acylated anthocyanins (i.e., CGG + FGG + SGG); Sum NAA = sum of non-acylated anthocyanins (i.e., G + GG) preceded by the type of root tissue (Ph = phloem, Xy = xylem), in the case of tissue-specific QTLs, and followed by "(%)" to indicate QTLs expressed as relative concentration. (**b**) Main characteristics of the segregating populations and purple-root sources used for mapping anthocyanin traits (Modified from Bannoud et al. (2019, 2021) and Iorizzo et al. (2020))

Besides the phenotypic loci described above conditioning the accumulation of anthocyanins in different tissues of the root and the aerial part of the plant, a major-effect trait locus controlling the proportion (%) of AA (relative to the total anthocyanin content) was discovered and mapped to Chr. 3, located between the P_1 and P_3 loci, at a distance of 17.9 cM from the latter (Cavagnaro et al. 2014) (Fig. 2a). This locus was termed *Raa1*, for "root anthocyanin acylation" (Cavagnaro et al. 2014). More recently, Curaba et al. (2020) performed fine mapping of *Raa1* in F_3 populations derived from the 70,349 family used by Cavagnaro et al. (2014) and an F_2 family obtained by crossing a 70,349-F_3 plant homozygous recessive at the *Raa1* locus with a homozygous dominant plant of Chinese origin known as "Ping Ding." Overall, the phenotyping and genotyping of a total of 926 plants allowed narrowing down the *Raa1*-containing genomic region, which was used for identifying candidate genes for *Raa1* (described in Sect. 2.3.3).

2.3.2 QTL Mapping

A first study concerning carrot anthocyanin QTLs was published by Cavagnaro et al. (2014) and reported on the genetic mapping of QTLs for individual and total root anthocyanins using an F_2 population derived from the Turkish purple-root progenitor P9547 (Fig. 2b). Fifteen significant QTLs conditioning total root anthocyanin levels, referred to as *RTPE* (for "root total pigment estimate"), and relative content of four cyanidin glycosides were identified and genetically mapped to five chromosomes. The QTLs with largest phenotypic effects (26.6–73.3% explanatory power) co-localized in two regions of Chr. 3. Five of these major-effect QTLs co-localized with P_3, and included a QTL for *RTPE* with 50% phenotypic effect, and four major QTL for concentration of three acylated pigments and one NAA (Fig. 2a). These results confirmed that the P_3 region controls anthocyanin accumulation in the carrot root and petioles in the P9547 genetic background (Cavagnaro et al. 2014). In a more recent study, fine mapping was performed in the P_3 region using a larger mapping population of the same genetic background as used earlier by Cavagnaro et al., detecting the same 5 major-effect QTLs for *RTPE* and four root cyanidin glycosides. Because of the larger population used, a better map resolution was obtained for this region, with overlapping QTLs for *RTPE*, Cy3XG, Cy3XSGG, and Cy3XFGG within a 3 cM region.

Bannoud et al. (2019) used two biparental populations (3242 and 5171) to map QTLs for concentration of individual and combined (i.e., AA, NAA, and total) anthocyanins, and relative percentages of individual and combined anthocyanins, in the root phloem and xylem. In the 3242 population, which was the most informative population with regards to the number of tissue-specific segregating traits, 41 and 8 QTLs for phloem and xylem anthocyanins, respectively, with QTLs explanatory power ranging from 1.3% to 53.4% were mapped across four carrot chromosomes (chromosomes 3, 4, 6, and 7). Thirty of the QTLs for phloem anthocyanins, including those with greatest phenotypic effect, and all of the QTLs for xylem anthocyanins co-localized to two region of Chr. 3, corresponding to the P_1 and P_3 regions previously described and mapped (Fig. 2a). While the QTL for phloem anthocyanins mapped to both of these regions, all the xylem QTLs mapped

to the P_3 region only, co-localizing with other phloem QTLs and the simply inherited traits *XAP*, *PAP*, *ROPAP*, and *RIPAP* conditioning presence/absence of pigmentation in the root xylem and phloem, and in leaf petioles. Altogether, these data strongly suggest that the P_3 region controls anthocyanin pigmentation (i.e., presence/absence of purple color) and concentration in the root phloem and xylem, and in leaf petioles, whereas P_1 conditions only phloem pigmentation and anthocyanin concentration, particularly in the outer phloem or cortex (Fig. 2a). The QTLs mapped in other chromosomes had relatively low explanatory weight, suggesting little phenotypic effect and/or interactions with major loci of Chr.3, as was revealed by Bannoud et al. (2019).

This chapter focused on the most important simply inherited phenotypic traits and QTLs with strongest effects conditioning anthocyanin pigmentation in carrot, as well as on the genes underlying these traits. These correspond to the P_1, P_3/RTPE, and *Raa1* regions in chromosome 3 controlling the presence, concentration, and acylation of these pigments. Additional and detailed information concerning all of the anthocyanin-related QTLs mapped to date across different chromosomes and carrot genetic backgrounds can be found in a recent review by Iorizzo et al. (2020).

2.3.3 Candidate Genes Conditioning Anthocyanin Biosynthesis, Acylation, Glycosylation, and Transport

By means of linkage analysis, Yildiz et al. (2013) attempted to identify candidate genes for P_1, mapping regulatory (*DcMYB3, DcMYB5,* and *DcEFR1*) and structural biosynthetic genes (*UFGT*, *PAL3, LDOX2*, *F3H*, and *FLS1*) of the anthocyanin pathway in the same population used for mapping the phenotypic trait. However, the map position of these genes did not coincide with P_1, indicating that none of them is a likely candidate for P_1. Later, Xu et al. (2014) compared gene expression of 13 anthocyanin structural genes in purple and non-purple roots of nice cultivars, finding upregulation in anthocyanin-containing roots for 9 of these genes, being their expression levels correlated with total root anthocyanin contents. The correlated expression of these anthocyanin genes with root pigment levels suggests a coordinated transcriptional regulation and their involvement in anthocyanin biosynthesis. However, no candidates for P_1, P_3, or *Raa1* – the phenotypic traits described prior to their study – could be identified.

A glycosyltransferase gene, termed *DcUCGalT1*, was isolated and cloned from the root of a purple carrot (Xu et al. 2016). The cloned gene encodes a galactosyltransferase that catalyzes the glycosylation of cyanidin with galactose. The enzyme has high affinity for galactosylation of cyanidin, but not for other anthocyanidins and flavonoid substrates, as indicated by its much lower galactosyltransferase activity for pelargonidin, peonidin, quercetin, and kaempferol observed in a heterologous expression system of *E. coli*. It was also found that the enzyme encoded by *DcUCGalT1* catalyzes exclusively the transfer of UDP-galactose to cyanidin, but it was unable to use other glycosyl donors like UDP-glucose or UDP-xylose, suggesting that *DcUCGalT1* is highly specific for galactosylation of cyanidin. Gene expression analysis of *DcUCGalT1* in nine purple and non-purple rooted cultivars revealed significantly greater expression in all purple roots compared to

non-purple ones, and the expression level was correlated with pigment content in the taproot. Together, these data indicate that *DcUCGalT1* is involved in cyanidin galactosylation. A BLAST search of *DcUCGalT1* in the carrot genome assembly showed greatest sequence homology with DCAR_009912 which is located in Chr. 3 at a position (12,350,188-12,351,646) unrelated to P_1, P_3, *Raa1*, *RTPE*, and the other tissue-specific phenotypic loci positionally associated with P_1 and P_3, suggesting that *DcUCGalT1* is not a major gene controlling anthocyanin pigmentation in the carrot root or petioles.

Following a similar approach as used for the *DcUCGalT1* gene, Xu et al. (2017) isolated and described an R2R3-type MYB transcription factor, designated *DcMYB6*, from purple carrot roots. The expression pattern of *DcMYB6* in purple and non-purple carrots was correlated with anthocyanin content in the roots. *Arabidopsis thaliana* plants genetically transformed with *DcMYB6* under a strong promoter showed anthocyanin pigmentation in various plant parts, including reproductive organs, paralleled with increased transcription of structural anthocyanin genes. These results suggest that *DcMYB6* upregulated the expression of Arabidopsis structural anthocyanin genes and may – presumably – play a similar role in the carrot root. In a subsequent study by Kodama et al. (2018), the transcriptomes of purple versus non-purple tissues from carrot roots and calli were compared by mean of RNA-Seq analysis and searched for differentially expressed members of the MYB, bHLH, and WD40 transcription factor families. They found a total of 104 differentially expressed TF genes, of which 32 had expression levels significantly correlated with root anthocyanin content. Additional analyses in other genetic stocks and root developmental times revealed that 11 of the 32 genes were consistently up- or downregulated in purple-rooted carrots, and included 6 MYBs, 4 bHLH, and 1 WD40 TF. Only one of these genes is located in Chr. 3, with the remaining ones being located throughout seven other chromosomes (and therefore are disqualified as potential candidates of the phenotypic traits in Chr. 3P_1, P_3/*RTPE*, and *Raa1*). The gene in Chr. 3 is a MYB TF (LOC108213488), which the authors referred to as *DcMYB6*, the same MYB gene described earlier by Xu et al. (2017). However, a closer analysis of the position of this gene in the carrot genome sequence revealed that the gene annotated as LOC108213488 is not *DcMYB6* but*DcMYB7*, another MYB TF described later as a potential candidate of P_3 by Iorizzo et al. (2019) and Xu et al. (2019).

Along with the sequencing and publication of the carrot genome, a list of 97 annotated structural biosynthetic flavonoid genes, including genes involved in anthocyanidin/anthocyanin biosynthesis and modification (e.g., glycosyltransferases and methyltransferases) was reported by Iorizzo et al. (2016). Besides facilitating this first large-scale catalogue of predicted structural genes involved in carrot flavonoid/anthocyanin biosynthesis, the availability of an annotated carrot genome sequence was instrumental for the discovery of key genes conditioning some of the phenotypic traits previously described and mapped. Among the first studies to take advantage of such an unprecedented genomic tool were those of Iorizzo et al. (2019), Xu et al. (2019), and Bannoud et al. (2019); the first two aiming at identifying candidate genes for P_3, while the latter investigated candidates for tissue-specific pigmentation in the root phloem and xylem.

Iorizzo et al. (2019) performed high-resolution mapping in the P_3/*RTPE* region of Chr. 3 using five mapping populations of different genetic backgrounds, including a larger-size population of the 70,349 F_2 used to first map these traits by Cavagnaro et al. (2014), for a total of 1669 plants phenotyped (421 individuals were phenotyped for *RTPE*, by HPLC analysis, and 1248 for P_3, by visual phenotyping of root color) and genotyped using different molecular marker types. The inclusion of common sequence-based SNP markers across the different maps allowed their comparative analysis to further delimit the region of interest, as well as anchoring such region of interest to the carrot genome assembly. Thus, by this approach, it was able to obtain a higher resolution of this map region, restricting the confidence interval of the *RTPE* QTL to a 2.6 cM region, whereas the closest flanking markers of P_3 delimited a 0.3–0.8 cM map region. A close examination of the linkage blocks harboring *RTPE* and P_3 in the different maps allowed further narrowing down the position of P_3 to a 494 kb region flanked by SNP markers K2590 and K0363 (Fig. 3). In this region, 8 anthocyanin-related transcription factor (TF) genes (6 MYB and 2 bHLH) were identified, whereas no structural anthocyanin biosynthetic genes were found. By means of gene prediction, orthologous, and phylogenetic analyses using MYB and bHLH TFs from other species with known functions, allowed the identification of six MYBs that had high sequence homology with MYBs involved anthocyanin biosynthesis, which were considered potential candidates for P_3/*RTPE.* The identified MYB TFs positionally associated with P_3/*RTPE* were designated as *DcMYB6*-*DcMYB11*, and they belong to the R2-R3-MYB family (Table 2). Noteworthy, *DcMYB6* corresponds to the MYB gene previously described by Xu et al. (2017). In the same study, the transcriptomes of purple versus non-purple roots and petioles were compared by means of RNA-Seq analysis, followed by validation of the candidate genes by RT-qPCR analysis. Of all six MYB TFs identified in the P_3 region, *DcMYB7* was the only gene consistently upregulated in all of the anthocyanin-containing tissues (as compared to non-purple tissues), suggesting that this MYB controls anthocyanin pigmentation in both the carrot root and petioles. *DcMYB11* was consistently upregulated in purple petioles (relative to green petioles) but not in purple roots, suggesting that it may co-regulate petiole pigmentation along with *DcMYB7*. Together, these results strongly point at *DcMYB7* as the main candidate for P_3 and a key gene conditioning root and petiole pigmentation, while *DcMYB11* specifically (co)regulates petiole pigmentation. These results by Iorizzo et al. (2019) suggest that, for the purple carrot lines used in their study, *DcMYB6* described earlier by Xu et al. (2017) does not play a major role in regulating anthocyanin biosynthesis in neither the taproot nor in the leaf petioles. Table 2 presents the MYB transcription factors associated with carrot anthocyanin pigmentation described to date, including their chromosome positions and gene IDs. The six anthocyanin-related MYB TFs found by Iorizzo et al. (2019) in the P_3 region are arranged in a small cluster of genes within a ~166 Kb region of Chr. 3 (Fig. 3b). A number of studies concerning genome wide analysis of R2-R3-MYB family members have revealed that MYB TFs are commonly found in gene clusters in the genomes of many plants species. Interestingly, it has been proposed that the expansion of MYB gene-families and their organization into gene clusters is mainly driven

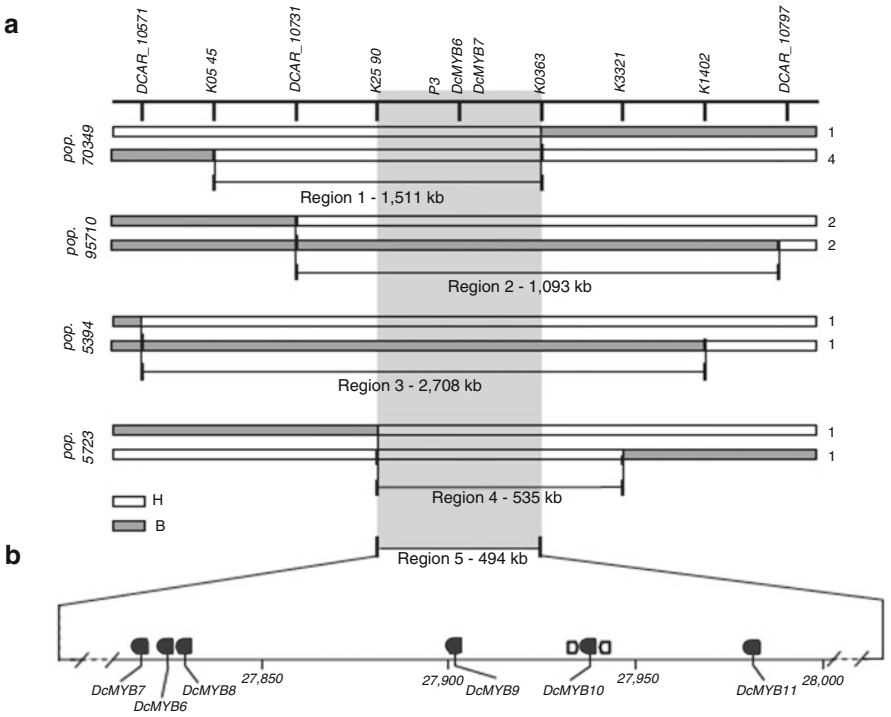


Fig. 3 Scheme of the fine mapping approach and identification of candidate genes in the P_3/*RTPE* region of chromosome 3 associated with anthocyanin pigmentation in the carrot root and petioles. (**a**) haplotypes delimiting the genomic regions controlling the *RTPE* QTL in population 70,349 (region 1), and the P_3 locus in populations 95,710, 5394 and 5723 (regions 2–4). White bars indicate the heterozygous haplotypes (*Aa*) and gray bars indicate the homozygous recessive haplotypes (*aa*). Region 5 represents the genomic sequence delimited by the nearest markers flanking *RTPE* and the P_3 locus across regions 1–4. The number of plants with recombinant genotypes is indicated by on the right of each bar. (**b**) Schematic representation of carrot chromosome 3 containing regions 1–5 and the six anthocyanin related MYBs (*DcMYB6-DcMYB11*) denoted in black boxes (Modified from Iorizzo et al. (2019))

by the occurrence of tandem and segmental duplications (Feller et al. 2011). Thus, it is possible that tandem and segmental duplications could have also influenced the genome organization of the carrot R2-R3-MYB gene family.

Xu et al. (2019) functionally characterized*DcMYB6* and *DcMYB7*, both candidates of P_3, by upregulating and shutting down their expression in carrot using a transgenic approach. It was found that overexpression of *DcMYB7*, but not *DcMYB6*, in the orange-rooted carrot cultivar 'Kurodagosun' led to the accumulation of large amounts of anthocyanin in the root and other tissues of this cultivar. Conversely, knockout of *DcMYB7* in the solid purple-rooted cultivar 'Deep Purple', obtained by gene-editing using the CRISPR-Cas9 system, resulted in carrots with yellow roots. These results confirm the role of *DcMYB7* as a major regulator of anthocyanin pigmentation in carrot root, as well as the main candidate of P_3. Their study also

Table 2 MYB transcription factors associated with carrot anthocyanin pigmentation described to date

Gene ID	Gene ID	Chr. #	Genome coordinates		References
			Start	End	
DcMYB1	DCAR_030745	9	29,370,465	29,370,597	Maeda et al. (2005)
DcMYB2	DCAR_011083	3	32,098,199	32,100,603	Wako et al. (2010), Meng et al. (2020)
DcMYB3-1	DCAR_028315	8	8,778,266	8,779,336	Wako et al. (2010)
DcMYB3-2	DCAR_028315	8	8,778,266	8,779,336	Wako et al. (2010)
DcMYB4	DCAR_015002	4	16,426,186	16,428,272	Wako et al. (2010), Kodama et al. (2018)
DcMYB4.2	DCAR_015002	5	22,226,323	22,229,683	Wako et al. (2010), Kodama et al. (2018)
DcMYB4.3	DCAR_028315	2	21,943,650	21,943,310	Wako et al. (2010), Kodama et al. (2018)
DcMYB5	DCAR_024737	7	18,692,086	18,693,681	Wako et al. (2010), Meng et al. (2020)
DcMYB6[a]	DCAR_000385	3	27,831,723	27,833,545	Xu et al. (2017), Iorizzo et al. (2019), Bannoud et al. (2019, 2021), Meng et al. (2020)
DcMYB7	DCAR_010745	3	27,816,911	27,819,103	Kodama et al. (2018), Iorizzo et al. (2019), Xu et al. (2019), Bannoud et al. (2019, 2021), Curaba et al. (2020), Meng et al. (2020)
DcMYB8	DCAR_010746	3	27,824,309	27,826,050	Iorizzo et al. (2019)
DcMYB9	DCAR_010747	3	27,901,372	27,903,024	Iorizzo et al. (2019)
DcMYB10	DCAR_010749	3	27,938,999	27,939,453	Iorizzo et al. (2019)
DcMYB11	DCAR_010751	3	27,980,959	27,982,962	Iorizzo et al. (2019)
DcMYB12/ DcMYB113	DCAR_008994	3	3,370,872	3,376,028	Bannoud et al. (2019, 2021), Xu et al. (2020)
DcMYB13	DCAR_009089	3	4,318,404	4,319,725	Bannoud et al. (2019)
DcMYB14	DCAR_010791	3	28,667,859	28,669,181	Bannoud et al. (2019)
DcMYB15	DCAR_010853	3	29,535,747	29,537,844	Bannoud et al. (2019)
DcMYB16	DCAR_003738	1	41,577,479	41,578,680	Meng et al. (2020)
DcMYB17	DCAR_007287	2	33,010,740	33,013,213	Kodama et al. (2018), Meng et al. (2020)
DcMYB18	DCAR_014214	4	23,846,982	23,848,124	Meng et al. (2020)
DcMYB19	DCAR_015602	4	10,021,890	10,023,032	Meng et al. (2020)
DcMYB20	DCAR_016459	5	3,566,349	3,567,632	Meng et al. (2020)
DcMYB21	DCAR_018481	5	30,483,119	30,484,602	Meng et al. (2020)
DcMYB22	DCAR_018882	5	34,067,212	34,068,530	Meng et al. (2020)
DcMYB23	DCAR_019908	6	35,858,709	35,860,309	Meng et al. (2020)
DcMYB24	DCAR_020645	6	29,610,878	29,613,179	Meng et al. (2020)

(continued)

Table 2 (continued)

Gene ID	Gene ID	Chr. #	Genome coordinates		References
			Start	End	
DcMYB25	DCAR_028146	8	14,422,378	14,423,688	Meng et al. (2020)
DcMYB26	DCAR_030321	9	23,408,024	23,410,706	Meng et al. (2020)
MYB1R1-1	DCAR_031036	9	32,191,963	32,196,585	Meng et al. (2020)
MYB1R1-2	DCAR_026095	7	33,127,347	33,129,362	Meng et al. (2020)
DcMYB27[b]	DCAR_026095	1	41,375,308	41,396,955	Bannoud et al. (2021)
DcMYB28[b]	DCAR_026095	8	9,013,410	9,014,567	Bannoud et al. (2021)

[a]*DcMYB6* was first described by Xu et al. (2017) and later incorporated manually into the carrot genome assembly [i.e., this gene was not included in the published genome assembly (Iorizzo et al. 2016)] with the coordinates above (Iorizzo et al. 2019)

[b]In the study of Bannoud et al. (2021), these MYBs were referenced with a generic name (i.e., "MYB" followed by their DCAR number); here we propose *DcMYB27* and *DcMYB28* to designate DCAR_026095 and DCAR_026095, respectively

demonstrated that *DcMYB7* could interact with – and activate the expression of – *DcbHLH3* (a homolog of the anthocyanin related *bHLH3* from *Malus domestica*) and structural genes of the anthocyanin biosynthetic pathway, and could also activate the expression of a putative glycosyltransferase (*DcUCGXT1*) and an acyltransferase (*DcSAT1*) genes identified in the same study (by orthologous analysis with related genes from other species). Together, these results provide strong evidence on the role of *DcMYB7* as a transcriptional activator of other regulatory and structural anthocyanin genes, including genes involved in pigment decoration (i.e., glycosylation and acylation). Interestingly, in non-purple carrots, the promoter of *DcMYB7* was interrupted either by *DcMYB8*, a nonfunctional tandem duplication of *DcMYB7*, or by two transposons, leading to the transcriptional inactivation of *DcMYB7*, thereby resulting in no anthocyanin synthesis and accumulation in their roots.

In the P_1 region, another MYB transcription factor, initially identified by Bannoud et al. (2019) and termed *DcMYB12*, was recently renamed as *DcMYB113* and functionally characterized by Xu et al. (2020). *DcMYB113* is an R2R3-MYB TF that controls anthocyanin pigmentation in the root of the carrot cultivar 'Purple Haze' and related genetic backgrounds [e.g., in B7262 (Bannoud et al. 2021)]. The function of *DcMYB113* is root- and tissue-specific, conditioning pigmentation only in the root periderm and outer phloem (also known as "cortex"). Also, as mentioned above, it is genotype-dependent, as indicated by the fact that this gene was not expressed in the roots of genetic backgrounds exhibiting anthocyanin pigmentation in both roots and petioles, including the commercial cultivars Deep Purple, Cosmic Purple, and Pusa Asita, and the inbred line P9547 (Xu et al. 2020; Bannoud et al. 2021). The root-specific activity of this gene was verified by genetic transformation of the orange carrot cultivar 'Kurodagosun' with *DcMYB113* fused to its own promoter and – in an independent transformation event – under the action of the CaMV 35S promoter. Transgenic 'Kurodagosun' plants carrying *DcMYB113* driven by the CaMV 35S promoter had solid purple roots and petioles, while the transgenic 'Kurodagosun' expressing *DcMYB113* with its own promoter had purple root and

green petioles, suggesting that root-specific expression of *DcMYB113* was determined by its promoter (Xu et al. 2020). As found previously for *DcMYB7*, it was also reported that *DcMYB113* could transcriptionally activate the expression of *DcbHLH3* and a number of structural anthocyanin biosynthetic genes, as well as two genes involved in pigment glycosylation (*DcUCGXT1*) and acylation (*DcSAT1*). Together, results from this study indicate that *DcMYB113* is the most likely candidate for P_1 and provide insights into its role in regulating anthocyanin biosynthesis and modification in the carrot root phloem.

Recently, a candidate gene for the *Raa1* locus, conditioning the relative content (%) of AA in the root, was identified and characterized by Curaba et al. (2020). By means of fine mapping in populations segregating for high AA versus low AA (individuals with more than 60% of the NAA Cy3XGG were scored as "low AA" and those with less than 22% of Cy3XGG were scored as "high AA"), followed by linkage blocks analysis, a genomic region of 576 kb harboring *Raa1* was identified and further searched for candidate genes. In this region, three predicted "Serine Carboxypeptidase-Like" (SCPL) genes, designated as *DcSCPL1-DcSCPL3*, were identified and analyzed at various levels. Phylogenetic analysis clustered the three carrot SCPLs with anthocyanin-related acyltransferase SCPL genes from other species, suggesting a similar role in carrot. Comparative transcriptome analysis indicated that only *DcSCPL1* was always expressed in association with anthocyanin pigmentation in the root and was co-expressed with *DcMYB7*. Structural (gene sequence) and expression analyses of *DcSCPL1* in roots with high AA and low AA revealed an insertion causing an abnormal splicing of the third exon during mRNA editing, likely resulting in the production of a nonfunctional acyltransferase and explaining the reduced acylation phenotype in the "low AA" plants. Together, results from this study strongly suggest that *DcSCPL1* is a major regulator of anthocyanin acylation in the carrot root and the main candidate of *Raa1*. Furthermore, the 700 bp-insertion found in the recessive allele of *DcSCPL1* was used to design flanking primers to develop a codominant PCR-based marker for aiding in the selection of plants with high or low AA in carrot breeding programs. It must be noted that the *DcSAT1* and *DcSCPL1* acyltransferase genes described independently by Xu et al. (2019, 2020) and Curaba et al. (2020), respectively, correspond to the same gene, with locus ID LOC108214129.

A few studies have addressed the genetics underlying tissue-specific pigmentation in the carrot root, using linkage mapping and comparative transcriptomic approaches in purple and non-purple root tissues. In addition to mapping QTLs for phloem and xylem anthocyanins, Bannoud et al. (2019) searched for candidate genes controlling pigmentation in the phloem, reporting that *DcMYB7* conditions both presence/absence and concentration of anthocyanins, while two cytochrome CYP450 genes with putative flavone synthase activity may negatively influence pigment content in this tissue. In a similar follow-up study concerning pigmentation in the outer phloem (*ROPAP*) and inner phloem (*RIPAP*), *DcMYB7* and *DcMYB113* appear as the main candidates for *ROPAP*, depending on the genetic background analyzed, while only *DcMYB7* conditions *RIPAP* (Bannoud et al. 2021). A MADS-box gene (DCAR_010757) located in the P_3 region was consistently upregulated in

association with purple pigmentation in the outer phloem (as compared to the non-purple inner phloem) in the two genetic backgrounds analyzed (3242 and B7262), as well as with *RIPAP* in P9547. However, this MADS-box was not phylogenetically related to MADS-box genes from other species known to be involved in anthocyanin biosynthesis and, therefore, its role in carrot anthocyanin biosynthesis is still inconclusive. Considering that the most striking differences in purple pigmentation between carrot cultivars are between the root xylem and phloem, comparative transcriptome analysis in these tissues may uncover major genes influencing their pigmentation. Such comparative analysis was conducted by Meng et al. (2020), finding 10 MYBs (Table 2) and 14 bHLHs, including *DcbHLH3*, differentially expressed between these two tissues. Interestingly, four of these MYBs, including two that were predicted as MYB1R1-like TFs (*DcMYB1R1-1* and *DcMYB1R1-2*), were downregulated in both purple phloem and purple xylem tissues, suggesting that they may act as transcriptional repressors of anthocyanin structural genes, thereby negatively regulating anthocyanin biosynthesis. Their role in carrot anthocyanin pigmentation deserves further investigation, as MYB transcriptional repressors of anthocyanin biosynthetic genes have been reported in several plant species, including the ornamental *Mimulus lewisii* and grapevine.

The regulation of anthocyanin metabolism ends with their transport into the vacuole, a process which can involve glutathione S-transferases (GSTs) (reviewed by Sylvestre-Gonon et al. 2019), and members of the "multidrug and toxic compound extrusion" (MATE) (Gomez et al. 2009) and "ABC-C transporter" gene families (Behrens et al. 2019). *DcGST1* (DCAR_003401), which co-localized with a minor-effect QTL on Chr. 1 conditioning total root anthocyanins, was reported to be upregulated in two solid purple cultivars compared to orange-rooted ones, and its upregulation correlated with increased vacuolar anthocyanin accumulation in purple carrots (Meng et al. 2020). Very recently, Bannoud et al. (2021) evaluated the expression of *DcGST1*, *DcMATE1* (DCAR-031151), and a carrot ABC Transporter gene (DCAR-010639) located in the P_3 region of Chr. 3, in purple versus non-purple root phloem tissues in two unrelated genetic backgrounds, B7262 (controlled by *DcMYB113*) and 3242 (controlled by *DcMYB7*), reporting significant upregulation of *DcGST1*and *DcMATE1* – but not of ABC Transporter – in all the purple tissues. Both of the upregulated genes were co-expressed with *DcMYB113* (in B7262) and *DcMYB7* (in 3242). Interestingly, according to Xu et al. (2020), *DcMYB113* was also co-expressed with *DcMATE1* in purple tissues of the cultivar 'Purple Haze'. Altogether, results from these three studies strongly suggest that *DcMYB113* and *DcMYB7* transcriptionally activate *DcGST1*and *DcMATE1* in their respective genetic backgrounds. Such regulatory role of MTB TFs on anthocyanin transporter genes has been reported in other plant species (Zhang et al. 2014).

2.4 Perspectives and Implications for Breeding

In recent years, candidate genes for the main phenotypic traits (P_1, P_3/*RTPE*, and *Raa1*) and QTLs conditioning anthocyanin biosynthesis, glycosylation, and

acylation across different carrot tissues and genetic backgrounds, as well as genes involved in cellular transport and vacuolar accumulation of these pigments, have been discovered and functionally characterized, and their causal mutations described and – for one of these genes (*DcSCPL1*) – used for the development of a molecular marker to aid in marker-assisted breeding (MAB). These advances are encouraging for breeding programs aiming at improving anthocyanin content and pigment profile for different end purposes, such as fresh consumption carrots with increased content of bioavailable NAA and – consequently – greater nutritional value; or carrots with high concentration of chemically stable AA suitable for the food colorant industry; and/or introgressing such pigment profiles into carrot genetic backgrounds with other desirable agronomic and consumer-quality traits. Also, the fact that these pigments have a role, in the plant, of ameliorating abiotic stresses (Shirley 1996), suggest that increasing general anthocyanin levels may result in more resilient carrot cultivars, of relevance in the current climate change context. Moreover, not only total anthocyanin content but also pigment profile, particularly with regards to AA:NAA ratios, may be relevant for this purpose, as it has been shown in other crops, such as red grapes, that exposure to high temperature results in reduced total anthocyanin content but increased proportion of AA, mediated by an upregulation of acyltransferases, suggesting that pigment acylation may be a mechanism for attenuating high temperature-stress consequences by reducing anthocyanin degradation (de Rosas et al. 2022).

To pursue these goals more rapidly and effectively, it would be ideal for breeding programs to combine well-stablished classic breeding strategies with molecular tools that allow tagging gene variants associated with specific pigment phenotypes. An example of such type of molecular tool is the codominant PCR-based marker targeting the *DcSCPL1* gene which allows differentiation of "low acylation" and "high acylation" alleles, which could be very useful for assisting early selection of plants with high or low proportion of AA. However, such MAB strategy would require developing additional markers for other key genes conditioning anthocyanin concentration and accumulation in different purple carrot backgrounds (e.g., *DcMYB7* and *DcMYB113*), and evaluating their usefulness for predicting pigment phenotypes across diverse carrot germplasm. Considering that the efficacy of molecular markers for predicting a phenotype relies on how tightly linked they are to the mutation that effects the phenotype, among other factors, marker development should ideally target the causal mutation of these genes, as was done with the *DcSCPL1* gene.

Alternatively, transgenic approaches represent technologically effective strategies for developing carrots with high anthocyanin content and specific pigment profiles. For example, increased anthocyanin concentration could be achieved by transforming carrots with *DcMYB7* under a strong constitutive promoter (CaMV 35S), whereas manipulating pigment profiles for high or low ratios of AA/NAA may be done using a similar strategy for upregulating *DcSCPL1* or generating a *DcSCPL1*-knockout by editing this acyltransferase gene, respectively. These and other transgenic approaches have been experimentally successfully applied to carrot for a number of traits (reviewed by Baranski and Lukasiewicz 2019), including

anthocyanin pigmentation (Xu et al. 2019, 2020). However, consumer's perceptions of transgenic crops vary widely across geographical regions and may represent an important barrier for a general wide acceptance of carrot cultivars developed by these methods. However, it can be speculated that food colorants derived from transgenic carrots may be accepted with less resistance than transgenic carrots for fresh consumption, especially considering the (publically perceived?) negative health effects associated with the consumption of some synthetic colorants.

3 Other Non-anthocyanin Phenolics

3.1 Introduction and Biosynthesis

As described for anthocyanins, phenolic compounds are secondary metabolites composed of an aromatic ring bearing one or more hydroxyl groups, and they exert similar functions in the plant as described for their colored-relatives anthocyanins, conferring protection against biotic and abiotic stresses (e.g., drought, salinity, ultraviolet radiation, extreme temperatures, phytopathogenic and predator attacks, etc.) and contributing to the organoleptic properties of plant foods (reviewed by Balasundram et al. 2006). Phenolic compounds can be subclassified as phenolic acids, flavonoids, tannins, lignans, stilbenoids, and curcuminoids. Carrot roots accumulate mainly phenolic acids, among which HCAs large predominate, being chlorogenic acid the major carrot HCA (Sharma et al. 2012). These phenolic acids are formed from the phenylpropanoid pathway, a late branch of the shikimic acid pathway. Figure 4 illustrates the core part of the pathway leading to HCAs and chlorogenic acid synthesis. HCAs are produced from phenylalanine through a series of reactions catalyzed by several enzymes, the first one being phenylalanine ammonia lyase (PAL), which converts phenylalanine to cinnamic acid. Cinnamic acid is then converted to coumaric acid by "cinnamate 4-hydroxylase" (C4H), and then to 4-coumaroyl CoA by "4-coumarate: CoA ligase" (4CL). 4-coumaric acid and 4-coumaroyl CoA represent important branch points in the phenylpropanoid pathway from which other subclasses of phenylpropanoids can be produced, such as coumarins, lignins, flavonoids, and hydroxybenzoic acids (another type of phenolic acids). 4-coumaroyl CoA is then converted to 4-caffeoyl quinic acid, known as chlorogenic acid, by a series of reactions catalyzed by the enzymes "hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase" (HCT), "hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferase" (HQT), and "4-coumarate-3-hydroxylase" (C3H) (Fig. 4).

3.2 Chemical Diversity and Distribution

Total phenolics content (TPC) in the carrot root varies broadly across accessions and root color phenotypes, as measured spectrophotometrically (Singleton et al. 1999; Fukumoto and Mazza 2000), with estimated ranges in different germplasm collections of 19.8–342.2 mg/100 g fw (Leja et al. 2013), ~25.9–426 mg/100 g fw

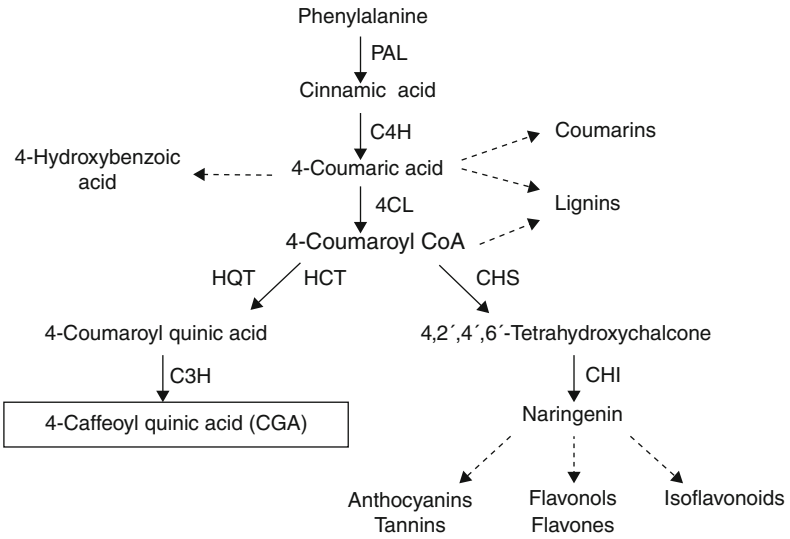


Fig. 4 Simplified phenylpropanoid biosynthetic pathway leading to the production of chlorogenic acid. Abbreviations used: *PAL* phenylalanine ammonia-lyase, *C4H* Cinnamate 4-hydroxylase, *4CL* 4-coumarate: CoA ligase, *HCT* hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase, *HQT* hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferase, *C3H* 4-coumarate3-hydroxylase, *CHS* chalcone synthase, *CHI* chalcone isomerase, *CGA* chlorogenic acid (4-caffeoylquinic acid) (Reproduced from Bartley et al. (2016))

(considering a dry matter content of 11% for expressing values on a fresh weight basis) (Sun et al. 2009), and 8.7–74.6 mg/100 g fw (Alasalvar et al. 2001). Because anthocyanins account for a large proportion of the TPC in purple carrots, the latter had much higher TPC than other root colors, with overall TPC means for purple carrots being 9 times (Leja et al. 2013) and 13 times higher than mean TPC in other carrot colors (Sun et al. 2009). In non-purple carrots, TPC levels varied little among the different root colors, with reported overall ranges of means – for white, yellow, orange, and red carrot cultivars – of 18–31, 24.2–40.4, and 18.3–25.9 mg/100 g fw, with mean differences among these root colors being statistically insignificant or marginally significant (Sun et al. 2009; Leja et al. 2013). By means of HPLC analysis, it was possible to precisely quantify non-anthocyanin TPC in purple and non-purple carrot cultivars, revealing significantly greater mean values in purple carrot (74.6 mg/100 g fw) than in orange (16.2 mg/100 g fw), yellow (7.7 mg/100 g fw), and white (8.7 mg/100 g fw) carrots (Alasalvar et al. 2001). Similar results were obtained in another HPLC-based study of phenolic composition in 10 carrot cultivars with different root colors grown across two locations and 3 years, reporting greater non-anthocyanin TPC in purple carrots (with a range of ~28.6–101.2 mg/100 g fw) than red (~ 0.07–6.41 mg/100 g/fw), orange (~ 0.09–3.41 mg/100 g fw), yellow (~ 0–0.93 mg/100 g fw), and white carrots (~ 0–0.77 mg/100 g fw) (for comparison purposes, values were converted on a basis of fw, considering 11% of dry matter for all the cultivars) (Kramer et al. 2012).

In the HPLC-based studies, non-anthocyanin TPC in different root color cultivars varied in the following rank order: purple > red > orange > yellow ≈ white (Alasalvar et al. 2001; Kramer et al. 2012). Altogether, data from these studies indicate that purple carrots generally have not only greater TPC but also greater total concentration of colorless phenolics than carrots of other colors.

Among the non-anthocyanin phenolics, HCAs and derivatives largely predominate in all the root color phenotypes (Kramer et al. 2012), including typical orange carrots where HCAs accounted for 73.7–99.7% of the root TPC (Zhang and Hamauzu 2004). With a few exceptions, in most of the cultivars analyzed to date, chlorogenic acid (5-*O*-transcaffeoylquinic acid) is the predominant HCA and – more generally – the predominant phenolic compound found in carrot roots. According to Kramer et al. (2012), this compound was detected in highest quantities in purple carrots, accounting for 54.1–79.7% of the total non-anthocyanin phenolics, while it showed broader ranges of variation in the other root colors, representing 5.1–64.9%, 13.2–68.2%, and 0–68.7% in red, orange, and yellow carrots, respectively, as varying across different growing locations and years. White carrots were the exception, with chlorogenic acid accounting for only 0.1–12.1% of the total phenolics, with vanillic acid derivatives being the predominant phenolics found in their roots, representing 10.4–83.6% of the TPC. Also, in one of the yellow cultivars evaluated (Yellowstone), ferulic acid and vanillic acid derivatives predominated over chlorogenic acid (Kramer et al. 2012). According to Alasalvar et al. (2001), chlorogenic acid represented 51–72% of the non-anthocyanin TPC in carrots of different root color (purple, orange, yellow, and white), whereas Zhang and Hamauzu (2004) reported that this compound accounted for 42–62% of the TPC in two orange-rooted cultivars. Besides the generally predominant chlorogenic acid, other phenolic acids that can be found in sufficient quantities in some carrot genetic backgrounds are derivatives of ferulic, vanillic, and caffeic acids (Alasalvar et al. 2001; Kramer et al. 2012).

In orange carrots, the concentration of phenolics varies across root tissues in the following order: periderm > phloem > xylem. According to Zhang and Hamauzu (2004), the root periderm only accounts for 11% of the carrot fresh weight, yet it provides 54.1% of the TPC in the root, whereas the phloem and xylem tissues provide 39.5% and 6.4% of the root TPC, respectively. Coincidently, chlorogenic acid levels and AOC in these three tissues followed the same relative ranking (i.e., periderm > phloem > xylem), with the periderm providing 67.4–80.1% of the total AOC and 66.5–88.2% of the total chlorogenic acid content, the phloem 7.5–9.6% and 10.7–29.6%, and the xylem 3.2–7.5% and 1.1–4.0%, for these two variables, respectively. These data illustrate the tissue-specific distribution of phenolic compounds in the carrot root and strongly suggest that chlorogenic acid is the major phenolic and antioxidant agent in orange carrots.

In line with the proposed role of phenolics in conferring protection to the plant against biotic and abiotic stresses, carrot TPC, chlorogenic acid, and AOC have been shown to concomitantly increase in response to numerous postharvest stimuli/stresses, including ultraviolet light radiation and wounding (Surjadinata et al. 2017), UVC light and hyperoxia (Formica-Oliveira et al. 2016), storage in modified

atmosphere (Pace et al. 2020), and water loss and wounding (Becerra-Moreno et al. 2015). In the field crop, nitrogen fertilization, boron deficiency, the cultivation system (organic vs. conventional), and the growing location and year can also influence carrot phenolic levels and AOC (Søltoft et al. 2010; Kramer et al. 2012; Singh et al. 2012).

3.3 Carrot Phenolics and Human Health

The consumption of polyphenols-rich plant foods has been associated with various health benefits, including maintenance of normal blood glucose and cholesterol levels, prevention or delayed onset of cognitive decline, and decreased risk of cardiovascular disease, diabetes, neurodegenerative disorders, and some cancer types (reviewed by Soto-Vaca et al. 2012). These general health-enhancing effects have been attributed – to a considerable extent – to the antioxidant and anti-inflammatory properties of phenolic compounds. Based on data concerning the relative impact on health of different phenolic subclasses, mainly with regard to their antioxidant, anti-inflammatory, and antiproliferative effects, the average dietary intake of polyphenols was estimated to be 1058 mg per day for males and 780 mg for females, with half of these composed of HCAs, 20–25% of flavonoids, and 1% anthocyanins (Stevenson and Hurst 2007).

A number of health-related studies, conducted in vitro and in vivo, have used purple carrot extracts as sources of anthocyanins and other phenolics, evidencing significant and substantial antioxidant, anti-inflammatory, and antiproliferative effects (described in Sect. 2.2.3). In germplasm evaluations including carrots of diverse geographical origins, root color, and genetic structure, significant and strong positive correlations have been found between TPC and AOC (evaluated by different methods), with reported correlation coefficient (r) values of 0.98–0.99 (for all root colors included) and 0.64–0.82 (considering only non-purple carrots) (Leja et al. 2013), 0.99 (includes all root colors) (Sun et al. 2009), 0.98–0.99 (includes only orange carrots) (Zhang and Hamauzu 2004), 0.62–0.93 (includes only orange carrots) (Surjadinata and Cisneros-Zevallos 2017), and 0.92 (only orange carrots) (Alegria et al. 2016). Similarly, significant positive and strong correlations between chlorogenic acid content and AOC were found, with estimated r values of 0.91 (Sun et al. 2009) and 0.89 (Alegria et al. 2016). Altogether, these data, and the fact that chlorogenic acid is the major carrot phenolic compound, strongly suggest that phenolics in general, and chlorogenic acid in particular, are largely responsible for carrot antioxidant properties.

Bioavailability is a relevant aspect for phenolic compounds being able to exert their health benefit effects in the organism. Generally, there is limited absorption despite the epidemiological studies showing significant health benefits associated with polyphenolic consumption. According to Soto-Vaca et al. (2012), the absorption of phenolic compounds in humans is a multifactorial, complex, and highly variable phenomenon, influenced by the type of phenolic compound, its molecular structure, the accompanying food matrix, interactions with other macromolecules and minerals (e.g., forming iron-chelating complexes), the processing by the

microbiome and resulting profile of phenolic metabolites, among many other factors. Because phenolics are generally considered to have poor bioavailability, there is considerable amount of literature – and ongoing studies – on these factors affecting the absorption of phenolic compounds (reviewed by Soto-Vaca et al. 2012).

3.4 Genetics of Carrot Phenolics

Compared to other nutraceutical classes, very little is known about the genetic regulation of non-anthocyanin carrot phenolics. Thus, very few studies have investigated possible genes involved in carrot HCAs and chlorogenic acid synthesis. A first work by Becerra-Moreno et al. (2015) examined the effect of water stress and wounding, applied alone or combined, on the expression of structural genes of the early shikimic acid pathway and later phenylpropanoid pathway, estimated by RT-qPCR analysis, along with the accumulation of related phenolic metabolites. Their findings indicate that both pathways are activated by these stresses applied alone, being wounding a stronger activator than water stress, while both stresses combined acted synergistically and showed the strongest pathway activation. Water stress favored the lignification process, while wounding led to the preferential accumulation of shikimic acid, phenolic compounds, and lignin. The increase in these phenolic compounds due to both types of stresses combined was accompanied by upregulation of PAL, C4H, 4CL, "caffeoyl-CoA3-O-methyl transferase" (CCoAOMT), "cinnamoyl-CoA reductase" (CCR), and "cinnamyl alcohol dehydrogenase" (CAD) genes involved in the phenylpropanoid pathway, whereas three upregulated genes of the early shikimic pathway were found, namely "3-deoxy-D-arabino heptulosonate 7-phosphate synthase" (DAHP synthase), 5-enolpyrovylshikimate 3-phosphate synthase (EPSP synthase), and chorismate mutase-prephenate dehydratase (CMPD).

Another study by Bartley et al. (2016) analyzed the expression of 12 structural genes (4 PAL, 2 CHS, and 1 each of C4H, 4Cl, HCT, HQT, C3H, and CHI) and 7 transcription factors (4 MYBs and 1 each of HY5, ERF, and UVR8) described or predicted earlier as being involved in phenylpropanoid synthesis in carrot and other species, in carrot root slices irradiated and not irradiated with UV-B light followed by incubation at 15 °C and 45% relative humidity for 6 days. Increases in gene expression – varying in intensity, duration of the peak maximum, and time of onset from the application of the UV treatment – after UV radiation were reported for all the structural genes except CHI, and most of the regulatory genes analyzed [except for ERF, which was not induced, and one of the MYB TF (termed *DcPcMYB1*) showing very little induction by UV] relative to their UV-untreated counterparts. These increases in gene expression by the UV treatment were paralleled with increases in the content of chlorogenic, caffeic, ferulic, and coumaric acids, compared to the content in untreated samples. These results suggest that most of the structural and regulatory genes evaluated in their study are involved in the UVB-induced synthesis of carrot HCAs.

While these two studies revealed genes involved in postharvest stress-induced synthesis of phenolic compounds, additional comparative studies using genetic

backgrounds varying in phenolic concentration and/or profile are necessary to identify genes influencing carrot phenolic levels under non-stressed field-cultivation conditions. Furthermore, in order to identify hierarchical genes controlling root phenolic content and profile, such studies should be combined with, or preceded by, linkage or association mapping approaches to localize the genomic region of the trait locus or QTLs conditioning such compositional variations, to then search for candidate genes – in the region of interest – using comparative gene expression analyses in contrasting phenotypes. This approach has been successfully used for identifying key genes controlling the type and content of other carrot nutraceuticals, namely anthocyanins and carotenoids (discussed in Sects. 2 and 6, respectively).

4 Polyacetylenes

4.1 Introduction

Polyacetylenes are a large and diverse chemical class of bisacetylenic oxylipins derived from fatty acids, characterized by the presence of at least two, usually conjugated, triple carbon-carbon bonds. Polyacetylenes are widely distributed in the plant kingdom, but they are also present in fungi, lichens, moss, marine algae, and invertebrates. The majority of the more than 2000 known polyacetylenes have been isolated from higher plants and notably from the botanically related plant families Apiaceae, Araliaceae, and Asteraceae (Dawid et al. 2015). The predominant structural types of polyacetylenes are different between Apiaceae and Asteraceae. In the Apiaceae and the closely related Araliaceae family, C_{17}-polyacetylenes with a variable number of additional double bonds and hydroxyl functions are dominating, while Asteraceae family members accumulate structurally diverse polyacetylenes. The most common C_{17}-polyacetylenes are falcarinol (FaOH) and falcarindiol (FaDOH). They occur in the edible and nonedible parts of vegetables and herbs of the Apiaceae family including carrots, parsnip, fennel, celery, and parsley. Falcarinol-type polyacetylenes, also called as falcarins (Santos et al. 2022), are less common in other food species, although falcarinol, falcarindiol, and related C_{14}- and C_{15}-polyacetylenes have been described for Solanaceae species like tomatoes and aubergines, where they are known to function as phytoalexins (Christensen 2011). Excessive amounts of falcarinol-type polyacetylenes have often been reported to contribute to the undesirable bitter off-taste of carrots and their products such as juice or puree. Quantitative chemical analyses combined with sensory perception tests indicate that falcarindiol is highly correlated with bitterness, whereas falcarinol is not (Czepa and Hofmann 2004). Apart from that, several in vitro studies have confirmed that falcarinol is among the most cytotoxic polyacetylenes in Apiaceae vegetables and, according to the current state of knowledge, has a higher bioactivity than falcarindiol (Christensen 2020).

The accumulation of falcarins in external plant tissue layers such as the root periderm is also in accordance with their assumed roles in plant defense, either as protective agents present constitutively in plant tissues or induced in response to

pathogen attack. For instance, in carrots, falcarindiol could inhibit the growth of the fungal leaf blight pathogen *Alternaria dauci* during in vitro tests, and the polyacetylene levels increased in response to *A. dauci* infection (Lecomte et al. 2012). However, compared with the knowledge available for polyacetylene effects on human health, there is still much to be learned about the role of these compounds in plant defense. With regard to putative effects of polyacetylenes on human health, falcarins have demonstrated numerous interesting bioactivities including antimicrobial as well as anti-inflammatory, antiplatelet-aggregatory, neuritogenic, and serotonergic effects. In addition, the cytotoxicity of falcarin-type polyacetylenes on human cancer cells and their proposed anticancer effects indicate that these compounds may contribute to the nutraceutical effects of carrots (Christensen 2011).

Despite their interesting biological functions, the genetic basis of the structural diversity and function of falcarins is widely unknown. A better understanding of the genetics of falcarin production in carrot roots might support breeding of carrot cultivars with tailored polyacetylenes levels for food production or nutraceuticals. The following sections summarize the knowledge on the distribution of falcarins in carrots, their bioactivity, biosynthesis, genetics, and genomics. A perspective is given for future carrot breeding programs aimed at elevating the levels of these specific polyacetylenes in specific carrot cultivars with high human health potential.

4.2 Diversity, Quantification, and Distribution

Among the more than 1400 polyacetylenes known to date in higher plants, a subset of 14 structurally related bisacetylenic oxylipins has been identified in *Daucus carota.* The first polyacetylenes isolated from carrots were falcarinol, falcarindiol, and falcarindiol-3-acetate. Besides these quantitatively predominating polyacetylenes which are typically C_{17} compounds with conjugated triple bonds, nine additional C_{17} falcarins were identified in *D. carota* by Schmiech et al. (2009), and Busta et al. (2018) recently identified two additional carrot polyacetylenes, falcarintriol-8-acetate and falcarintriol-9-acetate. A subset of five falcarin-type polyacetylenes present in carrot in noteworthy concentrations is shown in Fig. 5. A joining molecular characteristic of falcarin structures is their conjugated carbon–carbon triple bond system. Using this feature as a reference point to compare diverse falcarin structures by atom-to-atom correspondence analysis, a falcarin structural similarity network was constructed for 80 representative falcarins from the four families Apiaceae, Asteraceae, Araliaceae, and Solanaceae (Santos et al. 2022). This network suggests that falcarin metabolism may have diverged in these lineages and might support future efforts to elucidate falcarin biosynthesis in specific taxa.

Quantification of falcarinol-type polyacetylenes in tissues of carrots and other plants is generally based on chromatographic methods such as several HPLC and GC-MS methods. For the isolation of falcarins from plant extracts a combination of column chromatography (CC) and preparative and semi-preparative high-performance liquid chromatography (HPLC) have been used for the isolation of relative large amounts of these polyacetylenes, allowing the analysis of their bioactivity, for

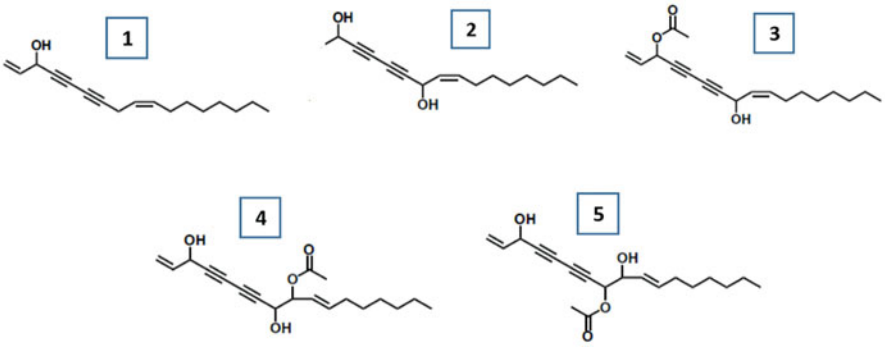


Fig. 5 Chemical structures of falcarin-type polyacetylenes present in carrot roots and leaves: (**1**) falcarinol, (**2**) falcarindiol, (**3**) falcarindiol-3-acetate, (**4**) falcarintriol-9-acetate, and (**5**) falcarintriol-8-acetate (Modified from Santos et al. (2022))

example, in preclinical studies. Modified qualitative and quantitative chromatographic methods have been described for falcarins, including analytical HPLC combined with UV-detection (Zidorn et al. 2005), HPLC with diode array detection (HPLC/DAD) (Christensen and Kreutzmann 2007), and capillary gas chromatographic techniques (GC–FID and/or GC–MS) (Czepa and Hofmann 2004). An alternative method based on plasma samples by liquid chromatography combined with mass spectrometry (LC–MS/MS) was used for the quantification of falcarinol and related polyacetylenes (Christensen and Brandt 2006). The nondestructive FT-Raman spectroscopic approach has been reported to differentiate the main falcarin compounds occurring in different tissues of a single plant (Krähmer et al. 2016).

In many reports, it has been shown that the concentration of falcarin-type polyacetylenes depends on factors such as genotype, developmental stage, cultivation, postharvest storage, and processing (for review, see Dawid et al. 2015). However, considering the numerous factors that influence the falcarin levels and the circumstances under which falcarin concentrations were obtained in the different studies, partly using different analytical methods, the falcarin levels published in different carrot papers are difficult to compare. Moreover, falcarin accumulation patterns in the carrot plant may be influenced by several abiotic and biotic stress factors. For instance, it has been shown that water stress due to drought or waterlogging may influence the major falcarin contents (Lund and White 1990), but previously it was reported that the genotypic effect (i.e., cultivar) appeared to have a stronger impact under these stress conditions (Schmid et al. 2021). The accumulation of falcarins can vary strongly among carrot cultivars, and major differences were found between cultivated orange carrots and differently colored carrots such as white, yellow, and purple cultivars (Czepa and Hofmann 2004; Schulz-Witte 2011). Recently, two main polyacetylenes, falcarinol and falcarindiol, were quantified by HPLC/DAD in different taproot tissues of seven carrot cultivars with purple, yellow, and orange taproot colors. Periderm tissue of the two purple cultivars, 'Deep Purple' and 'Anthonina', accumulated on average 580 or 720μg/g

dry weight (DW) falcarinol, respectively, but in the three orange cultivars significantly lower falcarinol levels ranging from 50 to 240μg/g DW were found. With regard to falcarindiol in root periderm, the purple cultivar 'Deep Purple' showed the highest concentration of 1230μg/g DW, whereas the levels of the orange cultivars were in the range of 450–650μg/g DW (Dunemann and Böttcher 2021).

The falcarin distribution in the carrot plant can vary considerably among the different plant organs and even within the different root segments. The most abundant falcarin in cultivated carrots is falcarindiol, and the highest total falcarin levels are generally found in the root periderm (Czepa and Hofmann 2004). Busta et al. (2018) analyzed tissue-specific accumulation of the five falcarins shown in Fig. 5 in the orange cultivar 'Danvers' and measured the highest total falcarin level in the periderm tissue, with falcarindiol as the dominating compound. In the purple cultivars 'Anthonina' and 'Deep Purple', the levels of falcarinol and falcarindiol were also significantly increased compared with phloem, xylem, and leaf tissue samples (Dunemann and Böttcher 2021). Detailed analysis of the spatial distribution of falcarins from the top to the bottom as well as from the outer phloem to the inner xylem of carrot roots indicate specific accumulation pattern dependent of the falcarin compound (Schmiech 2010). Results from in situ Raman spectroscopy experiments have indicated that polyacetylenes are located in vascular bundles in the young secondary phloem as well as in pericycle oil channels located close to the periderm layer. Moreover, analysis of the falcarin distribution in roots of some carrot wild relatives, for instance, *D. carota* ssp. *commutatus*, showed that the whole phloem tissue was enriched for falcarins with a maximum concentration near the pericyclic parenchyma (Baranska et al. 2005). Two novel falcarins detected previously in carrot cultivar 'Danvers' reached concentrations in the leaf, petiole, and root xylem, which are in the same magnitude as the levels of falcarinol, falcarindiol, and falcarindiol-3-acetate, with the difference that these novel falcarins appear to be accumulated preferably in the aboveground plant parts (Busta et al. 2018).

In the rare studies focused on polyacetylene accumulation in carrot wild relatives, it was shown that *Daucus* species and subspecies can contain often much higher falcarin contents than cultivated carrots (Schulz-Witte 2011). The concentration of falcarinol and falcarindiol in some wild *D. carota* accessions and other wild carrot (sub-) species, such as *D. c. maximus, D. c. maritimus,* or *D. c. halophilus* can be up to 10 or even 20 times higher in comparison to cultivated forms of carrots. Moreover, comparative quantification of falcarinol, falcarindiol, and falcarindiol-3-acetate in more than 100 accessions of wild carrot relatives revealed a large variation for falcarin content, indicating the enormous genotypic variability for this natural substance in the genus *Daucus.*

Falcarins such as falcarinol and falcarindiol can also be produced in vitro in large amounts by carrot hairy root cultures obtained after genetic transformation with the soil bacterium *Rhizobium rhizogenes.* HPLC/DAD quantification of falcarins in freeze-dried hairy root samples derived from different carrot cultivars revealed concentrations of these compounds that are considerably higher than reported to date for carrot root periderm tissue. The average falcarinol contents of the hairy roots varied from 1200 to 3000 μg/g DW, and the concentrations of falcarindiol were in the same magnitude (Dunemann and Böttcher 2021).

4.3 Bioactivity and Relevance for Human Health

Bioactive C_{17}-falcarins were first described for traditional medicinal plants such as ginseng (*Panax ginseng*, Araliaceae), the wild carrot Queen Anne's lace (*Daucus carota* subsp. *carota*, Apiaceae), and the understory native shrub Devil's Club (*Oplopanax horridus*, Araliaceae), all members of the closely related Apiaceae and Araliaceae families. Later it was found that the edible members of the Apiaceae family such as carrot, parsnip, celery, fennel, and parsley also contained falcarin-type polyacetylenes, and it was assumed that the content of these compounds, and falcarinol in particular, might be responsible for the beneficial effects of carrot consumption (Christensen and Brandt 2006). On the other hand, these phytochemicals might have some unwanted toxic side effects in plant foods. Some polyacetylenes, such as falcarinol, are powerful skin sensitizers and known to be neurotoxic in high concentrations, but at nontoxic concentrations, they may function as highly bioactive compounds with potential health-promoting properties. The number of reports about biological and pharmacological activities of polyacetylenes and in particular the contribution to health benefits associated with C_{17}-falcarins is increasing. In a recent review article, the cytotoxic, anti-inflammatory, and anticancer effects of C_{17}-falcarins and other acetylenic oxylipins from terrestrial plants including Apiaceae species are comprehensively described, and their possible mechanisms of action and structural requirements for optimal cytotoxicity are presented in detail (Christensen 2020).

With regard to anticancer actions of polyacetylenes in in vitro studies, both falcarinol and falcarindiol have been shown to have toxicity against a large variety of cancer cells including gastric, skin, intestine, colorectal, lymphoma, leukemia, breast, and lung, but their cytotoxic potential depends on the cell lines (Christensen 2020). Overall, from several in vitro studies it has been concluded that falcarinol is generally more cytotoxic than falcarindiol in many different cell types (Warner 2019). Furthermore, falcarinol and falcarindiol may have a synergistic inhibitory effect on cell proliferation. In the study of Purup et al. (2009), it was demonstrated that the cytotoxicity of lipophilic extracts from different carrot cultivars depended on the amounts of falcarinol, falcarindiol, and falcarindiol-3-acetate in the extracts. It was shown that falcarinol can reduce cell proliferation of Caco-2 cells at 2.5 μg/mL but maintains healthy cells at the same concentration. This was the case up to 10μg/mL at which concentration it significantly reduced proliferation of both cell types. Moreover, it was shown that the cytotoxic effect of falcarinol on Caco-2 cells was enhanced synergistically when combined with falcarindiol in different ratios. Falcarindiol was less potent in this experiment but reduced Caco-2 proliferation up to 20 μg/mL with no effect on healthy cells (Purup et al. 2009). Falcarinol was reported to have high cytotoxic activity to leukemia (L-1210), mouse fibroblast-derived tumor cells (L-929), mouse melanoma (B-16), and human gastric adenocarcinoma (MK-1) cells, showing the lowest ED_{50} values in the MK-1 cancer cells (Christensen 2011). Falcarindiol and its derivative falcarindiol-3-acetate isolated from carrots were able to induce apoptosis in different leukemia cell lines (CCRF-CEM, Jurkat, and MOLT-3). Falcarinol only caused induction of apoptosis

in CCRF-CEM cell lines but was the most cytotoxic substance on leukemia cells (Zaini et al. 2012). Falcarin-type polyacetylenes have also been shown to bind covalently to cysteine in enzymes such as mitochondrial aldehyde dehydrogenases (ALDHs) in cancer cells leading to a reduction of activity (Heydenreuter et al. 2015). Reduction of the activity of ALDHs may lead to oxidative stress and endoplasmic reticulum (ER) stress causing cell injuries, cell cycle arrest, and apoptosis, and thus could be one of the mechanisms of action that could explain the cytotoxicity of falcarins (Christensen 2020). Falcarins may also play a role in cancer chemotherapy. For instance, it has been demonstrated that falcarins function as inhibitors of the breast cancer resistance protein BCRP/ABCG2, which is an efflux transporter involved in breast cancer chemotherapy resistance (Tan et al. 2014). Bioactive C_{17}-polyacetylenes may also inhibit the formation of pro-inflammatory cytokines and inflammatory enzymes such as COXs and LOXs indicating that inhibition of these inflammatory-promoting substances might be an important mechanism of cancer prevention (Christensen 2020).

While there is strong evidence that polyacetylenes may have anticancer effects in vitro, much less information is available about the mechanisms and effects in vivo. A small number of investigations with rodents have been conducted with the aim to study both the effect of carrots and isolated falcarins as dietary supplements in feeding studies. However, little is known about their effects in vivo in humans. Rat feeding experiments with purified falcarinol and falcarindiol suggest preventing effects of these polyacetylenes on the development of colorectal cancer (CRC). Both falcarins can have an inhibitory effect on certain inflammatory markers in neoplastic lesions, and it has been proposed that falcarins may act as selective COX-2 inhibitors in relation to CRC prophylactics (Kobaek-Larsen et al. 2019). A large prospective cohort study in a Danish population of 57,053 individuals examining the risk of being diagnosed with CRC indicated a CRC-preventive effect of carrot intake. Self-reported intake of 2–4 raw carrots per week (>32 g/day) was associated with a 17% decrease in risk of CRC (Deding et al. 2020). These results and the results from the rat feeding experiments suggest a CRC-preventive effect of carrot falcarins.

Epidemiological studies have also shown that carrot intake is inversely associated with cardiovascular heart disease (CHD). In a prospective study, a 25 g/day increase in the intake of carrots was associated with a 32% lower risk of 10-year incidence of CHD (Oude Griep et al. 2011). One way the polyacetylenes could be affecting CHD risk is through modulation of platelet activity in the blood. Antiplatelet aggregatory abilities of polyacetylenes are most likely due to their anti-inflammatory activity in relation to COXs, LOXs, and other enzymes (Warner 2019). Furthermore, falcarins may also have impact on type 2 diabetes. Falcarinol and falcarindiol from carrots may function as ligands for nuclear receptors such as peroxisome-proliferator-activated receptors (PPARs), suggesting their usage as partial PPAR agonists with possible antidiabetic properties (El-Houri et al. 2015). Falcarinol was also identified as an inverse agonist of the cannabinoid receptor *CB1*. Acting by selectively alkylating the anandamide binding sites, falcarinol can increase the expression of chemokines in the skin involved in the induction of allergic reactions (Leonti et al. 2010). In addition, GABA receptors are exquisitely sensitive to polyacetylenic

oxylipins. In comparison to the activity of falcarindiol, falcarinol has a higher affinity for GABA receptors and a substantially different profile of pharmacological actions. Taken together, falcarin-type polyacetylenes are highly bioactive natural products that may be used in the prevention and treatment of cancers but likely possess additional positive properties with regard to human health and well-being. The health-promoting effects of carrot falcarins should be evaluated and confirmed in further preclinical and clinical studies in future.

4.4 Biosynthesis, Genetics, and Genomics

Compared with the extensive research concerning the analytical and biochemical identification and characterization of plant falcarin-type polyacetylenes, and the relatively large number of studies about their bioactivity, by far less is known about the biosynthesis, genetics, and genomics of falcarins. Major advances have only recently been made toward understanding their biosynthesis (reviewed by Santos et al. 2022), and carrots are among the species, where most knowledge exists. In carrots and other plants, falcarin biosynthesis starts from unsaturated fatty acids such as oleic acid and linoleic acid. Results of several metabolic studies have pointed out that a diverse pathway from linolenic acid to "unusual" fatty acids, such as crepenynic and dehydrocrepenynic acid, is the major route for the biosynthesis of falcarins (Minto and Blacklock 2008). A proposed model for falcarin biosynthesis is shown in Fig. 6.

The enzyme primarily responsible for the synthesis of linoleic acid from oleic acid is a Δ12-fatty acid desaturase. Numerous divergent forms of FATTY ACID DESATURASE 2 (FAD2) enzymes with diversified functionalities in fatty acid modification are also known to have diversified functionalities in fatty acid modification, such as hydroxylation, epoxidation, and the formation of acetylenic bonds and conjugated double bonds (Minto and Blacklock 2008). Some functionally divergent FAD2 enzymes are multifunctional, such as the bifunctional hydroxylase/desaturase from *Lesquerella fendleri.* FAD2 enzymes that introduce a triple bond within a fatty acid are designated as acetylenases. The first FAD2 acetylenase gene cloned was termed *Crep1* and derived from *Crepis alpina,* a herbaceous species from the Compositae family. *Crep1* encodes a functional acetylenase and its activity leads to the accumulation of the acetylenic fatty acid crepenynic acid in *C. alpina* seeds (Minto and Blacklock 2008). The genomes of other plant families like Apiaceae, Araliaceae, and Asteraceae also harbor FAD2 acetylenase genes, although only a few members of these families accumulate acetylenic fatty acids in their seeds, a fact that raises questions about the role of these divergent FAD2s in the plant.

The FAD2 gene family has been extensively studied in plants at the molecular and biochemical levels. Since cloning of the first plant FAD2 gene in *Arabidopsis thaliana* (Okuley et al. 1994), additional FAD2 genes from other species have been described and functionally characterized, including members with acetylenase activity. Although only a single FAD2 gene was found in Arabidopsis, in most other

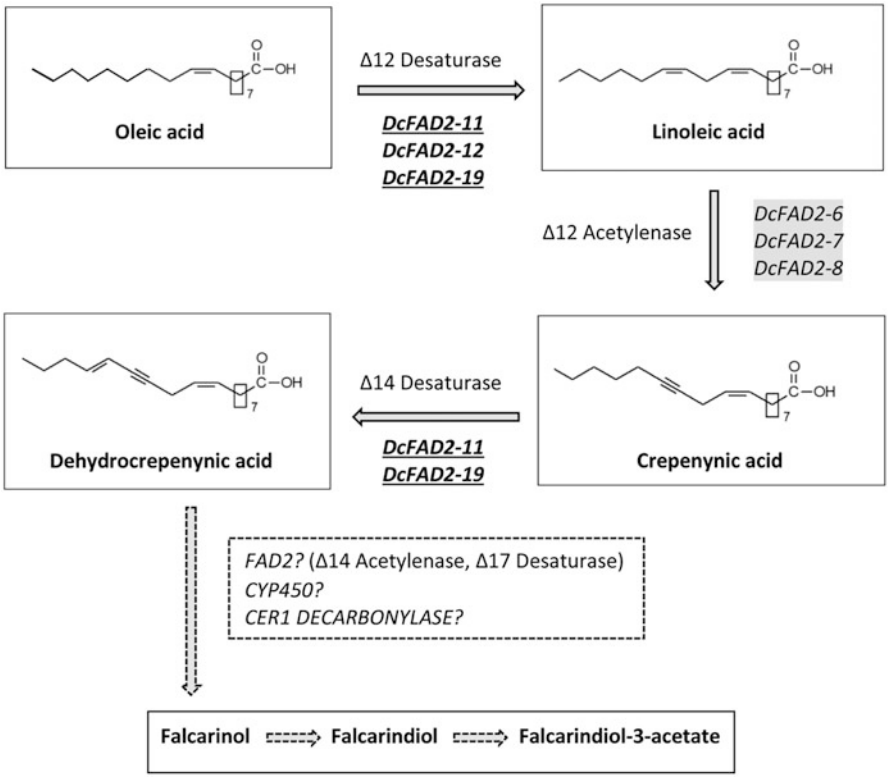


Fig. 6 Proposed biosynthetic steps of production of falcarin-type polyacetylenes in carrots. Carrot FAD2 genes functionally characterized by Busta et al. (2018) are shown at their presumed positions. FAD2 desaturase genes are in bold letters, and FAD2 acetylenase genes are shaded in gray. Bifunctional FAD2 desaturase genes are underlined. FAD2 gene designation was according to Table 4 (Adapted from Cavagnaro (2019) and Santos et al. (2022))

plant species, multiple FAD2 homologues have been reported. For example, seven FAD2 gene family members were identified in soybean, and nine each in cotton and tomato (Lee et al. 2020). Species with the largest known number of FAD2s are all family members of the Asteraceae, Apiaceae, and Araliaceae. The carrot genome contains at least 30 FAD2 genes (Dunemann et al. 2022). In sunflower (*Helianthus annuus*, Asteraceae), about the same number of FAD2 genes is present, and for ginseng (*Panax ginseng*, Araliaceae), over 50 FAD2 gene copies have been reported recently (Santos et al. 2022). A major contribution for the inventory and putative function of carrot FAD2 genes was made by the work of Busta et al. (2018), which identified 24 carrot FAD2s. To identify which of these genes participate in falcarin biosynthesis, six FAD2 genes were selected for functional analysis by screening of published, tissue-specific carrot transcriptomic data sets for co-expression of genes that are highly expressed in root periderm. These candidate genes were heterologously expressed, individually and in combination, in yeast and *A. thaliana*,

resulting in the identification of one canonical FAD2 that converts oleic to linoleic acid and three divergent FAD2-like acetylenases that convert linoleic into crepenynic acid (Busta et al. 2018). In addition, two bifunctional FAD2s with Δ12 and Δ14 desaturase activity were found (DCAR_013547, DCAR_019786) that are involved in two biosynthetic steps, the conversion of oleic acid into linoleic acid and the conversion of crepenynic into dehydrocrepenynic acid. Three of the characterized FAD2s are located in a tandem array of six FAD2genes, two of which encode FAD2 acetylenases and one encoding a bifunctional Δ12/Δ14 desaturase, in a small region (29.30–29.35 Mbp) of carrot chromosome 4 (Busta et al. 2018). Tandemly arrayed FAD2 genes, which might represent biosynthetic gene clusters, were previously described also in tomato, lettuce, and sunflower and might have been originated from local tandem duplications, as suggested from phylogenomic and microsyntenic analyses (Busta et al. 2018). The largest carrot FAD2 cluster in the carrot genome, consisting of nine FAD2s, is located on chromosome 8, spanning the genomic region 20.87–21.54 Mbp (Dunemann et al. 2022). This gene cluster also contains the newly discovered FAD2 genes *DcFAD2-25* to *DcFAD2-28*. The seven new carrot FAD2 genes identified recently after a reannotation of the whole carrot genome sequence extend the total number of *Daucus* FAD2s to 31. The total list of known carrot FAD2s and a proposed renaming are shown in Table 3. The newly predicted FAD2 genes *DcFAD2-29*, *DcFAD2-30*, and *DcFAD2-31* are located as unclustered genes on chromosomes 1, 3, and 4. A phylogenetic analysis of the carrot FAD2s (Fig. 7) showed that *DcFAD2-25*, *DcFAD2-26*, *DcFAD2-27*, and *DcFAD2-28* are located in a common clade containing ten known *Daucus* FAD2s which have been classified as divergent FAD2s. This clade contains also the genes *DcFAD2-6* (DCAR_017011), *DcFAD2-7* (DCAR_013552), and *DcFAD2-8* (DCAR_013548) which were functionally characterized by Busta et al. (2018) as Δ12-acetylenases. These three genes are also highly similar (amino acid identity >95%) to the parsley gene *PcELI12,* which was identified as a pathogen-inducible FAD2 acetylenase gene (Cahoon et al. 2003).

Table 3 List of 31 *Daucus carota* FAD2 gene models sorted by their physical position on the assembled 9 carrot chromosomes according to the whole-genome sequence (Iorizzo et al. 2016) and the carrot FAD2 inventory published by Busta et al. (2018) and Dunemann et al. (2022)

Chrom.	Gene name	Locus name	Genomic position Start	Stop	Protein lenght
1	***DcFAD2-29***	Not annotated	2623973	2625091	373
1	***DcFAD2-18***	DCAR_002026	24320363	24321512	382
1	***DcFAD2-3***	DCAR_003420	38720179	38721331	383
3	***DcFAD2-30***	Not annotated	11882922	11884034	371
3	***DcFAD2-14***	DCAR_011708	39497199	39498366	388
3	***DcFAD2-15***	DCAR_011709	39502342	39503509	388
4	***DcFAD2-31***	Not annotated	10151245	10152378	378
4	***DcFAD2-22***	DCAR_013553	29308974	29310153	392
4	***DcFAD2-7***	DCAR_013552	29312792	29313944	383
4	***DcFAD2-17***	DCAR_013551	29318510	29319686	391
4	***DcFAD2-16***	DCAR_013549	29340355	29341522	388
4	***DcFAD2-8***	DCAR_013548	29343013	29344165	383
4	***DcFAD2-19***	DCAR_013547	29348356	29349505	382
5	***DcFAD2-13***	DCAR_017010	10549274	10550426	383
5	***DcFAD2-6***	DCAR_017011	10559828	10560980	383
5	***DcFAD2-23***	DCAR_017012	10593543	10594722	392
5	***DcFAD2-24***	DCAR_017923	24911311	24912904	367
5	***DcFAD2-11***	DCAR_019786	41732457	41733609	383
5	***DcFAD2-20***	DCAR_019787	41738175	41739333	385
6	***DcFAD2-10***	DCAR_020161	33657256	33658405	382
6	***DcFAD2-12***	DCAR_019845	36316385	36317537	383
7	***DcFAD2-5***	DCAR_025967	32083676	32084828	383
8	***DcFAD2-4***	DCAR_027655	20875213	20876365	383
8	***DcFAD2-25***	Not annotated	20877925	20879073	383
8	***DcFAD2-26***	Not annotated	20881526	20882677	383
8	***DcFAD2-27***	Not annotated	20885064	20886215	383
8	***DcFAD2-28***	Not annotated	20887620	20888768	383
8	***DcFAD2-1***	DCAR_027616	21164452	21165604	383
8	***DcFAD2-9***	DCAR_027615	21168523	21169678	384
8	***DcFAD2-21***	DCAR_027614	21171805	21172963	385
8	***DcFAD2-2***	DCAR_027583	21538656	21539808	383

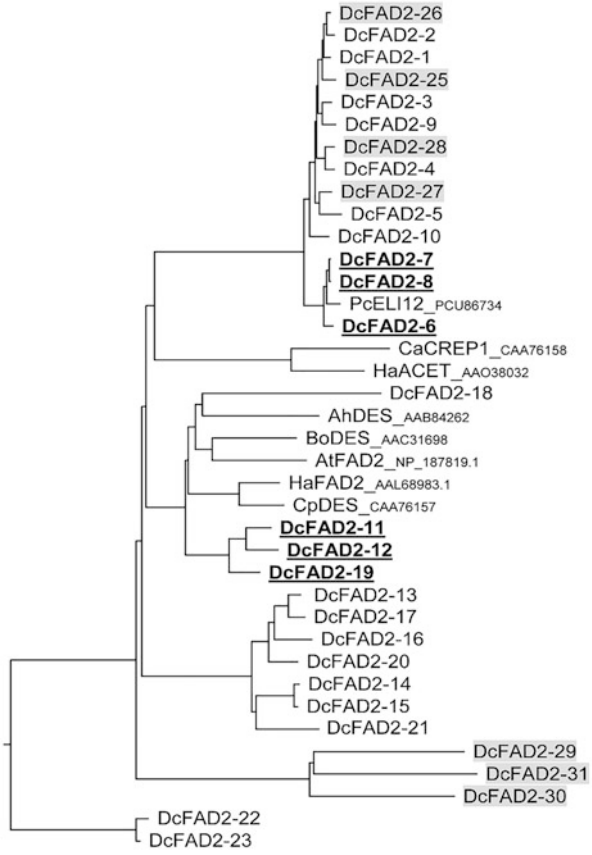


Fig. 7 Similarity of *Daucus* FAD2s shown by a phylogenetic tree (ClustalW) of deduced FAD2 protein sequences from *D. carota* (Busta et al. 2018, Dunemann et al. 2022) and some other putative FAD2 acetylenases and desaturases (Pc *Petroselinum crispum*; Ca *Crepis alpina*; Ha *Helianthus annuus;* At *Arabidopsis thaliana*; Bo *Borago officinalis*; Cp *Crepis palaestina*; Ah *Arachis hypogaea*). Carrot FAD2s newly identified by Dunemann et al. (2022) are shaded in gray. Carrot FAD2 genes with known function are indicated in bold/underlined (Busta et al. 2018)

The biochemical formation of falcarindiol from its presumed precursor falcarinol in later steps of the falcarin pathway was largely unknown until now. It is hypothesized that plants having the enzymatic machinery necessary for cuticular waxes should also be able to perform decarbonylation and in-chain hydroxylation of fatty acids required for falcarin biosynthesis (Santos et al. 2022). However, the substrates and mechanisms for these processes are unclear. Recently a pathogen-induced biosynthetic gene cluster was discovered in tomato and shown to be putatively involved in falcarindiol production (Jeon et al. 2020). The four genes of this cluster were consistently and strongly co-expressed, three of which were FAD2 acetylenases or desaturases, with the remaining one being a CER1 decarbonylase homologue. In *Arabidopsis thaliana*, CER1 (ECERIFERUM1) and CER3 (ECERIFERUM3) gene products catalyze fatty acid decarboxylation in the alkane biosynthetic pathway and play a major role in wax production (Aarts et al. 1995). Although the study of Jeon et al. (2020) implicates a possible role for CER1 in falcarindiol production, a functional proof for a possible involvement of CER1 in the decarbonylation of falcarin precursors has, to date, not been described.

Compared with the outcome from research on identification and functional characterization of genes putatively involved in biosynthesis of falcarin-type polyacetylenes, relatively less is known about the genetic control of falcarin contents in carrots. However, in carrot breeding programs aiming at reducing bitter taste (e.g., for baby food or juice and puree production), or – conversely – increasing pathogen resistance and/or enhanced health benefits, more information about the inheritance of falcarin accumulation is needed. Especially the availability of molecular markers linked to QTLs, or usage of functional markers developed directly from candidate TPS genes, would strongly support such breeding strategies. First QTLs for carrot falcarins were identified by using a segregating F_2 family and indicated that levels of falcarin compounds are heritable traits (Le Clerc et al. 2019). In the carrot population used in this study over a 2-years period for QTL identification of three major falcarins, a relatively low environmental influence on the falcarin contents was found, and broad-sense heritability, for example, falcarindiol, was estimated to be 0.88. A transgressive segregation pattern was also observed in this study. Several QTLs could be mapped for falcarindiol and falcarindiol-3-acetate in carrot roots, and some of these QTLs were also associated with QTLs for bitterness and resistance to the fungal leaf blight pathogen *Alternaria dauci* (Le Clerc et al. 2019). In another biparental carrot F_2 progeny derived from a cross of a cultivated breeding line and the wild relative *D. carota* subsp. *commutatus*, large phenotypic variability was obvious for the contents of falcarinol and falcarindiol (Dunemann et al. 2022). The observed frequency distributions in this F_2 family indicated a polygenic inheritance, which is expected considering the assumed complex biosynthetic pathway of falcarin production. Nevertheless, the large phenotypic variation, ranging from individuals with no measurable falcarins to plants with very high contents of 1000 μg/g DW (dry weight) falcarinol or even 3000 μg/g DW falcarindiol, suggests the action of major regulatory or structural genes. However, significant falcarin QTLs were identified on six of the nine carrot chromosomes, which is a sign for the complexity of the genetic control of falcarin production in carrot roots. Common QTLs with high LOD scores were identified for both falcarinol and falcarindiol on carrot chromosomes 4 and 9, indicating a major involvement of these two genomic regions in polyacetylene biosynthesis (Dunemann et al. 2022).

The six-gene FAD2 cluster associated with QTLs on chromosome 4 contains the two genes, *DcFAD2-7* and *DcFAD2-8*, which have been functionally characterized as Δ12-fatty acid acetylenases, and *DcFAD2-19*, described as a bifunctional desaturase (Busta et al. 2018). The genetic association of three CER1 candidate genes with the strongest QTLs on chromosome 9 might be a further indication that these genes might also play a role in falcarin production. However, a functional proof that CER1 genes contribute to the production of falcarin-type polyacetylenes is still missing and remains to be biochemically established.

4.5 Implications for Breeding

Falcarin-type polyacetylenes are among the phytochemicals that may contribute to bitter taste of fresh carrots or carrot products, being an undesirable trait that can

cause consumer rejection. Therefore, carrot breeding generally aims at low bitterness. Bitterness is a very complex trait, because not only polyacetylenes but also a variety of other chemically different potential bitter compounds, such as volatile mono- and sesquiterpenes, phenylpropanoids, and isocoumarins, may be involved in this quality trait. Recently, Schmid et al. (2021) analyzed and listed more than ten known bitter off-taste compounds present in carrot roots including the major falcarins, falcarinol and falcarindiol. Nevertheless, the falcarins appear to be a lead substance for bitter taste. In the study of Le Clerc et al. (2019), total falcarin content was closely related to bitterness, and the highest quantities of this compound accumulated in the most bitter genotypes, whereas the lowest amounts were measured in the least bitter genotypes. It is likely that lowering falcarin levels in carrots occurred in the past during the domestication process and breeding of cultivated orange carrot forms with the desired low bitterness. However, carrot cultivars with relatively high falcarin contents, such as the purple-rooted cultivar 'Anthonina' are still available in the seed market. This dark-purple cultivar also contained the highest contents of other putative health-relevant phytochemicals such as special phenylpropanoids and flavonols, when compared with white, yellow, orange, and red cultivars (Leja et al. 2013). As shown in a carrot biodiversity study, 'Anthonina' is closely related to carrot landraces originating from continental Asia (Baranski et al. 2012). Therefore, this cultivar is a good example that early domesticated purple carrots might be used directly as nutraceuticals or, alternatively, for the large-scale production of pharmaceutically relevant falcarins to be used, for instance, as food supplements. On the other hand, some modern purple F_1 hybrid cultivars, such as 'Deep Purple', might also be suited. Considering the proposed health-promoting effects of falcarins, specific carrot chemotypes with high but acceptable amounts of bioactive falcarins may be developed in future by molecular breeding using gene-specific functional markers or, to reach the goal faster, by application of genome editing techniques such as CRISPR/Cas. For both approaches, it will be necessary to first reveal the most important genes that control the decisive steps in the biosynthesis of falcarins.

5 Terpenes

5.1 Introduction

Terpenes are the largest and most structurally diverse class of plant natural products, comprising tens of thousands different substances. They can be classified into the subclasses monoterpenes (C_{10}), sesquiterpenes (C_{15}), diterpenes (C_{20}), and triterpenes (C_{30}), according to the number of isoprenoid structures. In plants, low-molecular-weight terpenes produced by terpene synthases (TPS) contribute to multiple ecologically and economically important traits. Low-molecular-weight terpenes are involved in plant-to-plant communication and plant protection against abiotic and biotic stresses. They play an important role in plant defense against insect, fungal, and bacterial pathogens, which has been demonstrated in numerous studies

(Yoshitomi et al. 2016; Chen et al. 2018; Shaltiel-Harpaz et al. 2021). A typical characteristic of mono- and sesquiterpenes is their volatility, and therefore they contribute to the typical flavor and aroma of many plant species including carrot. The typical flavor of carrots is determined by a complex blend of mono- and sesquiterpenes, representing up to more than 90% of the total volatile compounds. In a few studies, the correlation between volatile terpenes and carrot sensory attributes has been investigated (Alasalvar et al. 2001; Fukuda et al. 2013). Because carrots contain a huge amount of different terpene compounds, it is difficult to relate single terpene compounds to specific aroma notes and positive or negative sensory perception by humans. Using a GC olfactometry (GC-O) approach, an association between the carrot aroma and flavor and several isolated terpenes could be established (Kjeldsen et al. 2003). On the other side, terpenes are often involved in harsh or bitter flavor notes, and these off-flavor characteristics were shown to increase with terpene contents in different carrot genotypes. The combination of chemical analyses with sensorial approaches has predicted the monoterpenes sabinene, α-terpinolene, and β-pinene as candidates for bitterness in carrots (Le Clerc et al. 2019). Terpenes are present abundantly in essential plant oils of many plants, and they are also important components of resins and floral scents. Although there is general acceptance that volatile terpenes play a role as bioactive substances which might have an impact on human physiology and health, the current knowledge about putative health effects of terpenes present in carrots is restricted. Furthermore, the existing knowledge is rarely based on studies with terpenes isolated directly from *D. carota* but is generally founded on terpenoid compounds present in considerable quantities in other plant species such as, for instance, *Cannabis sativa* or the tea tree *Melaleuca alternifolia.*

A better understanding of the genetic and molecular bases of the TPS enzymes involved in terpene biosynthesis will be needed for future breeding strategies aimed at the improvement of quality traits, such as aroma and taste, but might be also relevant for phytopathological aspects, for example, resistance against fungi or insects. In addition, the development of specific carrot chemotypes to be used for pharmaceutical applications might be feasible. In the following sections, we briefly describe the analysis, the spectrum of volatile terpenes in carrot roots and seeds, and their putative bioactivity with regard to human health. We further review the current knowledge about TPS enzymes and genes involved in terpene biosynthesis in carrot.

5.2 Diversity, Quantification, and Distribution

The most common analytical method for determining volatile terpenes is headspace solid-phase microextraction gas chromatography (HS-SPME-GC). This method was used in several investigations on carrot roots, leaves, and seeds, either coupled with flame-ionization-detection (FID) or with mass-spectrometry (MS). As a sensitive sampling method, headspace (HS) sampling of terpenes emitted by carrot tissues has been proved successful to achieve a complete qualitative and quantitative volatile organic compound (VOC) analysis. Particularly, solid-phase micro-extraction

(SPME), which is a popular and simple HS sampling technique requiring no or little preparation of the samples, was mostly used in the past for the analysis of carrot VOCs, including terpenes. For example, using HS-SPME-GC-FID and MS, Ulrich et al. (2015) analyzed the diversity of terpene volatile patterns of carrot roots and leaves, and the same method was used to examine the terpene profiles in a large collection of carrot cultivars (Keilwagen et al. 2017). Since a full multi-compound quantitation using the SPME technique is impossible in complex organic matrices like homogenates of carrot leaves and roots, semiquantitative data are usually reported, for instance, as relative concentrations. Auto-HS-SPME-GC-MS was used for terpene analysis of differently colored carrot roots and carrot seed oil (Yahyaa et al. 2015, 2016). Headspace sorptive extraction (HSSE), a technique similar to SPME, was applied in carrots to link special volatile terpene profiles with sensory attributes (Fukuda et al. 2013). This technique uses a twister stir bar coated with polydimethylsiloxane (PDMS) to adsorb VOCs in the headspace of diced carrot tissue prior to thermal desorption and GC-MS analysis (Ibdah et al. 2019). For determination of absolute amounts of terpenes by the headspace sampling method, a precise calibration using authentic standards is required. An alternative is to use ground plant tissue for extraction with an organic solvent such as hexane and to use the extracts for GC-FID or GC-MS with the appropriate internal standards (Ibdah et al. 2019).

Detailed chemical analysis of the volatile terpene pattern in carrot leaves and roots revealed differences in total amounts and proportions of individual compounds, suggesting large tissue-dependent differences in terpene biosynthesis (Habegger and Schnitzler 2000). In contrast to carrot roots, the terpene profiles of leaves have rarely been studied. Conversely to results of Hampel et al. (2005), reporting no correlation between leaf and root terpenes, Ulrich et al. (2015) showed strong correlations for some compounds present in leaves and roots. Another investigation based on different carrot root tissues revealed that the biosynthesis of terpenes is mainly localized in the phloem (Hampel et al. 2005). Higher concentration of terpenoids in the root phloem than in the xylem has been explained by the observation that oil ducts, a possible site of volatile terpenoid biosynthesis, are found only in the phloem. Nevertheless, in the root xylem, biosynthesis of terpenes was also detectable, even in the absence of oil ducts in this tissue (Hampel et al. 2005).

The existing knowledge about volatile terpene patterns in carrot roots is mainly based on research aimed at the identification of compounds involved in flavor, aroma, and taste. The majority of the terpenes identified from roots are mono- and sesquiterpenes (Fig. 8). Although a comparatively large number of publications is available for VOCs in carrots, no accordance exists about typical carrot volatile profiles. The lack of agreement in finding typical patterns of major terpenes related to flavor and taste was documented by Ulrich et al. (2015), who examined about ten publications dealing with carrot terpenes. Out of the more than 120 compounds described in the literature, about 80 VOCs are single entries and were identified only in a single study. In 11 studies, the following 5 terpenes were identified: sabinene, limonene, terpinolene, β-caryophyllene, and α-humulene. In addition, other frequently mentioned volatiles were α-pinene, β-pinene, γ-terpinene, and p-cymene

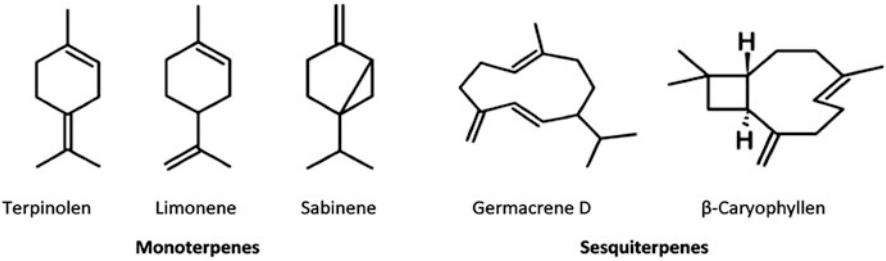


Fig. 8 Chemical structures of typical mono- and sesquiterpenes present in carrot roots, leaves, and seeds

(Ulrich et al. 2015). The monoterpenes α-terpinolene and α-pinene were found as the main terpenes in extracts from 11 differently colored carrot cultivars, and among the totally identified 16 monoterpenes, the monoterpenes α-pinene, β-pinene, β-myrcene, d-limonene, γ-terpinene, α-terpinolene, and p-cymene constituted more than 60% of total VOCs identified (Güler et al. 2015). A compilation of predominant mono- and sesquiterpenes identified by different researchers in fresh roots of differently colored carrots was shown by Ibdah et al. (2019). In the study of Keilwagen et al. (2017), 85 *D. carota* accessions representing the worldwide gene pool of cultivated carrots were analyzed for the qualitative and semiquantitative composition of VOCs in both roots and leaves. Totally, 31 VOCs, mostly monoterpenes and sesquiterpenes, were identified in roots or leaves, and some of them were present in both organs. In leaves, the most abundant compounds were β-myrcene, β-caryophyllene, and limonene, whereas terpinolene, β-caryophyllene, and bornyl acetate dominated the VOC patterns of roots. The majority of the root VOCs were monoterpenes, whereas only four sesquiterpenoids were present in the roots, namely β-caryophyllene, β-farnesene, β-bisabolene, and caryophyllene oxide. Five of the compounds (sabinene, β-caryophyllene, ocimene, α-pinene, and terpinen-4-ol) showed correlations between leaves and roots, which was in accordance with earlier results for sabinene and β-caryophyllene (Ulrich et al. 2015; Keilwagen et al. 2017). Twenty-three of these substances were also present in a biparental carrot F_2 mapping population consisting of 320 individuals (Dunemann et al. 2019). In the work of Koutouan et al. (2018), targeted analyses of terpene volatiles in carrot leaves identified compounds potentially linked to resistance to the leaf fungus *Alternaria dauci.* In total, 30 terpenes, 15 monoterpenes, and 15 sesquiterpenes were quantified over several environments and years. In all genotypes, the main monoterpenes were β-myrcene, sabinene, α-pinene, and limonene, and the main sesquiterpenes were caryophyllene and germacrene. Seven terpenes differentiated resistant genotypes from the susceptible H1 genotype. Compared with carrot cultivars, very little information is available about terpene distribution and their amounts in wild *Daucus* relatives. A rare example is the study of Reduron et al. (2019), in which seedling root extracts of wild populations of *D. carota* from Corsica (France) were chemically analyzed and separated by their individual VOC profiles. For instance, the *D. c.* ssp. *commutatus* population 786, adapted to stony soils and a dry and hot habitat, showed

very high contents of limonene and sabinene in roots, and γ-himalachene in leaves, indicating an eco-subspeciation process and adaptation to specific environmental conditions (Reduron et al. 2019).

The cultivated carrot is mainly used as a root vegetable, while its seed oil is sometimes employed as a flavoring agent in food products, in the cosmetics industry, and aroma therapy. Therefore, a number of investigations have focused on the identification and quantification of VOCs from essential oils of carrot seeds. Since wild carrots have been used as a medicinal plant since ancient times, the majority of analyses was performed in carrot wild relatives and local natural carrot populations. Aćimović et al. (2016) analyzed the essential oil from seeds of wild (*D. carota* ssp. *carota*) and cultivated carrots (*D. carota* ssp. *sativus*) collected in northern Serbia and reported that the oil derived from wild-grown carrots contained mainly sabinene and *α*-pinene, followed by β-bisabolene, β-pinene, and *trans*-caryophyllene as further dominant compounds. The major constituents of essential oil from seeds of the cultivated carrots were carotol, sabinene, and *α*-pinene. Sabinene and *α*-pinene, together with myrcene, p-cymene, and limonene, were also the major compounds among 70 terpenoid VOCs identified in nine carrot cultivars (Flamini et al. 2014). In another study based on seeds of wild growing *D. carota* ssp. *carota in* Lithuania, the oils from all samples were of the sabinene chemotype. The other major constituents were α-pinene, terpinen-4-ol, γ-terpinene, and limonene (Mockute and Nivinskiene 2004). In the work of Sieniawska et al. (2016), the chemical composition of commercially available (Moroccan and French) and hydrodistilled (Polish) wild carrot seed essential oils from *D. carota* ssp. *carota* was analyzed, and the sesquiterpene alcohol carotol was found to be the main constituent in three seed oils. Pinene, sabinene, myrcene, limonene, geranyl acetate, bisabolene, caryophyllene oxide, and daucol were identified as other main compounds. Profiling the terpene metabolome in carrot seeds of wild carrot accessions from Israel revealed significant differences in volatile composition among the accessions. In most, but not all the accessions, the monoterpene α-pinene was the most abundant volatile, followed by limonene (Yahyaa et al. 2016). As discussed by these authors, there are many factors, such as genotypic differences, stage of development, environment, and geographical origin that can considerably influence the volatile composition pattern of *D. carota* seed oil. It is therefore not surprising that values obtained in the different studies are inconsistent to some extent.

5.3 Bioactivity and Relevance for Human Health

Despite the large diversity of terpenoid volatiles in carrot roots and their high relevance for flavor and taste, their biological effects have been particularly associated with the essential oils of the carrot seed and their compounds, which seem to be primarily terpenes. Essential oil of *D. carota* ssp. *carota* from Portugal, enriched for geranyl acetate, α-pinene, and other terpenes showed antibacterial and antifungal activity against several Gram-positive and Gram-negative bacteria, yeasts, dermatophytes, and *Aspergillus* strains. In addition, this seed oil was also able to inhibit germ

tube formation and preformed biofilms of *Candida albicans* (Alves-Silva et al. 2016). Essential oil derived from *D. carota* ssp. *halophilus* has displayed antifungal properties against several human pathogenic fungi (Tavares et al. 2008), while the oil from *D. carota* ssp. *maritimus* has been shown to have potential antibacterial effects (Jabrane et al. 2009). Antimicrobial and antioxidant activities of carrot seed oil were also reported by Jasicka-Misiak et al. (2014). Vasudevan et al. (2006) described antinociceptive and anti-inflammatory properties of wild carrot seed extracts, and Shebaby et al. (2013) demonstrated antioxidant and anticancer effects. Hydrodistilled yellow and red carrot oils derived from seeds of *D. carota* ssp. *sativus* (yellow carrot) and D. c. *ssp. boissieri* (red carrot) were able to suppress 5-LOX and prostaglandin E2 production indicating excellent anti-inflammatory activities (Khalil et al. 2015). The cytotoxicity of both essential oils was evaluated on two human cancer cell lines, namely HepG-2 (liver hepatocellular carcinoma cells) and MCF-7 (breast adenocarcinoma cells), after 72 h incubation. The highest cytotoxic activity was observed against HepG-2 cells with IC_{50} values in the range of 163–172 μg/ml for both oils (Khalil et al. 2015). The observed cytotoxic activity of the two oils may be caused by carotol, which is the major component in both yellow and red carrot oils. Carotol was also among the main compounds in three carrot seed oils of different origin, amounting 19–33% of the sum of compounds (Sieniawska et al. 2016). Cytotoxicity tests based on green monkey kidney (VERO) and human pharynx squamous cell carcinoma (FaDu) cell lines treated with isolated carotol indicated nonselective moderate cytotoxicity on both cell lines, but apparently this substance was not solely responsible for the cytotoxic effects of the seed oils (Sieniawska et al. 2016). In another study, carotol was reported to inhibit the growth of myeloid leukemia cancer cell lines (Radoslaw 2012). The bioactivity of essential oil from wild carrot seeds is also probably due to its high contents of the monoterpenes sabinene and α-pinene. For instance, both terpenes are major components of pharmaceuticals which are used to treat the protozoan *Trypanosoma brucei* causing the African sleeping sickness disease.

Other reports about the wide spectrum of bioactive effects of carrot terpenes are available from studies with other plant species containing the same terpene compounds. For example, the monoterpene limonene, comprising up to >90% of orange peel oil, and other monoterpenes have shown chemopreventive activity against rat mammary, lung and forestomach cancers (Crowell 1999). The sesqiterpene β-caryophyllene (BCP) was shown to increase the cytotoxic activity of paclitaxel in various cancer cell lines (Legault and Pichette 2007), with the largest observed effect on DLD-1 cells treated with paclitaxel plus 10 μg/mL BCP. Application of geraniol has been shown to sensitize cancer cells to the conventional chemotherapeutic agent 5-fluorouracil (5-FU), and it supported an increased uptake of the drug (Carnesecchi et al. 2004). The monoterpene terpinen-4-ol is known as the main component of the oil of tea tree (*Melaleuca alternifolia*), which has pharmaceutical importance due to its known antibacterial and anticancer effects (Lee et al. 2020). The sesquiterpene BCP is not only one of the major terpenoids in carrots but also known as a major plant volatile present in large amounts in the essential oils of many different spice and food plants. BCP is a major component (up to 35%) in the

essential oil of *Cannabis sativa* and has been identified as a functional non-psychoactive *CB2* receptor ligand in foodstuff and as a macrocyclic anti-inflammatory cannabinoid (Gertsch et al. 2008). BCP selectively binds to the cannabinoid receptor *CB2* and functions as a *CB2* agonist. Among numerous cannabinoids, BCP has received attention in the past few years due to its therapeutic potential by mediating anti-inflammatory and immunomodulatory properties. The various pharmacological properties and the therapeutic potential of BCP are comprehensively summarized by Jha et al. (2021). Only recently, it has been hypothesized that BCP could be a promising therapeutic agent to target the triad of infection, immunity, and inflammation in COVID-19 (Jha et al. 2021). Evidence from a preclinical study also suggests that a BCP-*CB2* interaction might be involved in anxiety and depression disorders of mice, and that *CB2* receptors may provide alternative therapeutic targets for the treatment of anxiety and depression (Bahi et al. 2014). The monoterpenes β-pinene and linalool can directly hit the central nervous system and have an enhanced activity after inhaling (Guzmán-Gutiérrez et al. 2014). Another interesting finding is that β-pinene can interact with dopaminergic D1 receptors, which is a mechanism used by many antidepressant drugs (Cox-Georgian et al. 2019).

5.4 Biosynthesis, Genetics, and Genomics

Terpenoids represent the largest class of natural substances produced by land plants. Many of these secondary metabolites are highly specialized and often involved in interactions with the environment. Thus, continued innovation of terpenoid biosynthesis has played an important role in the adaptation of land plants during evolution and diversification. The large diversity of terpenes is formed by members of terpene synthase (TPS) family from few substrates by similar carbocation-based reaction mechanisms (Tholl 2006). Terpenoids derive from the isomeric 5-carbon building blocks isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). Two independent pathways in plants, the methylerythritol phosphate (MEP) pathway operating in plastids and the mevalonate (MVA) pathway operating in the cytosol/peroxisomes, lead to the synthesis of both IPP and DMAPP. Geranyl pyrophosphate (GPP) or geranylgeranyl pyrophosphate (GGPP) formed in the MEP pathway, and farnesyl pyrophosphate (FPP) formed in the MVA pathway are the substrates for terpene synthases, which catalyze the production of mono- or diterpenes from GPP or GGPP, respectively, or sesquiterpenes from FPP. Many volatile terpenes are formed directly by TPS enzymes, and various other enzyme classes are involved in modification of the primary terpene skeletons. These other enzymes belonging for a major part to the classes of dehydrogenases, methyl- and glycosyltransferases, and cytochrome P450 hydroxylases, can increase the volatility of terpenes and modify their olfactory features (Pateraki et al. 2015).

In plants, the classification of the TPS gene family comprises eight subfamilies, known as TPS-a to TPS-h, based on their protein sequences and functional characteristics (Chen et al. 2011). The subfamilies TPS-a, TPS-b, and TPS-g are only

present in angiosperms and, among these, the TPS-a subfamily includes mainly sesquiterpene and diterpene synthases, whereas members of the TPS-b subfamily are responsible for the production of monoterpenes or isoprenes. Members of the smaller TPS-g subfamily are predominantly monoterpene synthases. The TPS-d subfamily is only found in the gymnosperms, whereas TPS-h is specific to the spikemoss *Selaginella moellendorffii*. The TPS-c subfamily is, presumably, the ancestral clade, and it includes several copalyl diphosphate synthase genes. The TPS-f clade derived from the TPS-e subfamily, and they are, therefore, often combined and referred to as the TPS-e/f subfamily, which harbors ent-kaurene synthase genes and other diterpene, monoterpene, and sesquiterpene synthase genes (Chen et al. 2011).

Analysis of sequenced plant genomes have revealed varying TPS gene family sizes among species, ranging from 30 to more than 100 members, which presumably evolved – as for many other gene families – by duplication followed by functional divergence. For example, the genomes of *A. thaliana*, rice, and grapevine contain 32, 34, and 69 full length TPS, respectively (Chen et al. 2011). More than 100 putative TPS genes were identified each in the representative *Eucalyptus* species *E. grandis* and *E. globulus* (Külheim et al. 2015). TPS gene families have also been studied and described extensively in several other plant species such as tomato (Falara et al. 2011) and *Vitis vinifera* (Martin et al. 2010).

Using the first published carrot whole-genome assembly, Iorizzo et al. (2016) performed a preliminary characterization of the carrot TPS family and predicted 36 potentially functional TPS genes. Using the same version of the carrot genome, Keilwagen et al. (2017) conducted a genome-wide identification of carrot TPS genes by applying a newly developed homology-based gene prediction software called GeMoMa. In total, the carrot TPS inventory actually consists of 65 predicted full-length TPS gene models. The predicted carrot TPS genes were assigned, with one exception (TPS-d) to all angiosperm TPS subfamilies, TPS-a to TPS-g. Subfamily TPS-b (monoterpene synthases) was identified as the largest subfamily in the carrot genome with a total of 32 putatively functional genes. Subfamily TPS-a (sesquiterpene synthases) was found to be the second largest subfamily with 22 genes. The huge number of TPS-a and TPS-b genes correspond with the dominance of mono- and sesquiterpenes present in the large panel of carrot genotypes (Keilwagen et al. 2017). Results from this study revealed that carrot is among the species with high or very high number of TPS genes, suggesting broad potential for further functional diversification and specialization of terpene metabolism. Analyzing the carrot genome sequence, most of the TPS genes are found in gene clusters, dispersed throughout all of the chromosomes (Keilwagen et al. 2017). Especially the TPS gene clusters on chromosomes 1, 3, 4, and 9 might be results of multiple gene duplications (Fig. 9). The whole known inventory of *Daucus* TPS genes together with existing information about their putative function is shown in Table 4. A maximum-likelihood phylogenetic tree including all predicted TPS genes in *D. carota* was presented by Ibdah et al. (2019).

In contrast to the high impact of mono- and sesquiterpenes for the total volatile profile and, therefore, for taste and flavor of carrots, only a few research papers had

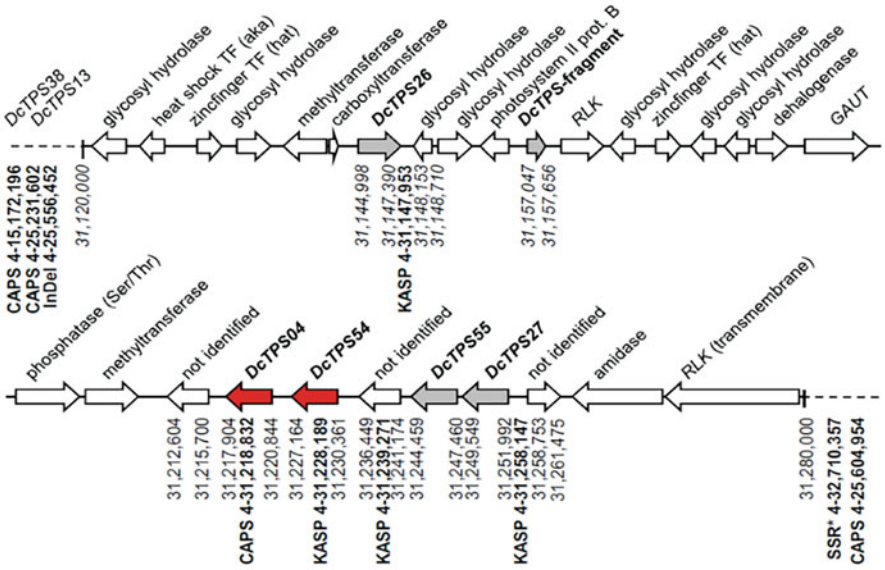


Fig. 9 Schematic representation of the physical location of the TPS gene cluster on chromosome 4 of the carrot genome (Reproduced from Reichardt et al. 2020)

focused on functional characterization of carrot TPS genes. A *DcTPS1* recombinant protein synthesized in a heterologous expression system in *E. coli* produced mainly the sesquiterpenes β-caryophyllene and α-humulene, while recombinant *DcTPS2* functioned in *E. coli* as a monoterpene synthase with geraniol as the main product (Yahyaa et al. 2015). *DcTPS2* is a gene encoding a monoterpene synthase of the TPS-b subfamily and produced also β-myrcene, which is among the carrot terpenes with strongest influence on carrot aroma. The function of *DcTPS2* as a geraniol synthase was confirmed by analysis of a very similar gene cloned from a wild carrot accession, which contained geraniol as the predominant terpene in seeds (Yahyaa et al. 2016). Based on QTLs mapping results and the genetic dissection of a five-TPS gene cluster on chromosome 4 with molecular markers, Reichardt et al. (2020) selected *DcTPS04* and *DcTPS54* for further functional analyses. In vitro enzyme assays in *E. coli* showed that *DcTPS54* encodes a single-product enzyme catalyzing the production of the monoterpene sabinene, whereas *DcTPS04* was found to be a multiple-product terpene synthase producing α-terpineol as a major product and four other products, namely sabinene, myrcene, β-pinene, and β-limonene. These results were confirmed by Muchlinski et al. (2020), who functionally characterized 19 carrot TPS genes in the double-haploid orange carrot genotype DH1 and compared spatial expression profiles and in vitro products of their recombinant proteins with the volatile terpene composition of DH1 and the respective terpene profiles from four genotypes with different root colors. As TPS enzymes involved in the production of major compounds of carrot flavor, *DcTPS07* and *DcTPS11* (germacrene D), *DcTPS30* (γ-terpinene), and *DcTPS03* (α-terpinolene) were biochemically identified

Table 4 List of 64 *Daucus carota* TPS gene models sorted by their physical position on the assembled 9 carrot chromosomes according to the whole-genome sequence (Iorizzo et al. 2016) and the carrot TPS inventory published by Keilwagen et al. (2017). For functionally characterized TPSs, only the main product is named together with the reference publication: (1) Yahyaa et al. (2015), (2) Reichardt et al. (2020), and (3) Muchlinski et al. (2020)

Chrom.	Gene name	Locus name	TPS subfamily	Main product (Reference)
1	***DcTPS32***	DCAR_002080	TPS-b	
1	***DcTPS11***	Not annotated	TPS-a	germacrene D (3)
1	***DcTPS45***	DCAR_002829	TPS-g	
1	***DcTPS46***	DCAR_002830	TPS-g	
1	***DcTPS19***	DCAR_002831	TPS-g	linalool (3)
1	***DcTPS47***	DCAR_004091	TPS-b	
1	***DcTPS10***	Not annotated	TPS-b	
1	***DcTPS24***	Not annotated	TPS-b	
1	***DcTPS48***	Not annotated	TPS-b	linalool (3)
1	***DcTPS22***	Not annotated	TPS-b	
1	***DcTPS49***	Not annotated	TPS-b	
2	***DcTPS41***	Not annotated	TPS-a	
2	***DcTPS40***	Not annotated	TPS-a	
2	***DcTPS42***	Not annotated	TPS-a	germacrene D (3)
2	***DcTPS03***	Not annotated	TPS-b	α-terpinolen (3)
3	***DcTPS15***	Not annotated	TPS-a	α-phellandrene (3)
3	***DcTPS50***	Not annotated	TPS-a	
3	***DcTPS37***	Not annotated	TPS-a	
3	***DcTPS08***	Not annotated	TPS-a	
3	***DcTPS51***	Not annotated	TPS-b	
3	***DcTPS05***	Not annotated	TPS-b	
3	***DcTPS12***	Not annotated	TPS-b	
3	***DcTPS18***	Not annotated	TPS-b	
3	***DcTPS25***	DCAR_012483	TPS-c	
3	***DcTPS31***	Not annotated	TPS-b	
3	***DcTPS52***	DCAR_012537	TPS-b	
3	***DcTPS30***	DCAR_012538	TPS-b	γ-terpinene (3)
3	***DcTPS53***	Not annotated	TPS-a	δ-elemene (3)
3	***DcTPS06***	Not annotated	TPS-a	
4	***DcTPS38***	Not annotated	TPS-a	
4	***DcTPS13***	Not annotated	TPS-a	
4	***DcTPS26***	DCAR_013310	TPS-b	limonene (3)
4	***DcTPS04***	DCAR_013298	TPS-b	α-terpineol (2)
4	***DcTPS54***	DCAR_013297	TPS-b	sabinene (2), (3)
4	***DcTPS55***	DCAR_013294	TPS-b	sabinene (3)
4	***DcTPS27***	DCAR_013293	TPS-b	
4	***DcTPS09***	DCAR_012965	TPS-b	
4	***DcTPS02***	DCAR_012963	TPS-b	geraniol (1)
5	***DcTPS56***	DCAR_016843	TPS-e	
5	***DcTPS28***	DCAR_016844	TPS-e	
5	***DcTPS14***	DCAR_017536	TPS-b	
5	***DcTPS17***	DCAR_018214	TPS-b	
5	***DcTPS57***	DCAR_018422	TPS-c	
5	***DcTPS58***	DCAR_019208	TPS-b	
5	***DcTPS33***	DCAR_019208	TPS-b	
5	***DcTPS59***	DCAR_019490	TPS-c	
6	***DcTPS01***	DCAR_023152	TPS-a	β-caryophyllene (1)
7	***DcTPS23***	DCAR_024752	TPS-g	
7	***DcTPS60***	DCAR_024753	TPS-g	
7	***DcTPS61***	Not annotated	TPS-a	
8	***DcTPS62***	DCAR_028138	TPS-b	
8	***DcTPS21***	Not annotated	TPS-b	
8	***DcTPS29***	DCAR_027915	TPS-f	
8	***DcTPS44***	DCAR_026972	TPS-b	
8	***DcTPS43***	DCAR_026971	TPS-b	
9	***DcTPS36***	Not annotated	TPS-a	
9	***DcTPS35***	Not annotated	TPS-a	
9	***DcTPS63***	Not annotated	TPS-a	
9	***DcTPS64***	Not annotated	TPS-a	
9	***DcTPS07***	Not annotated	TPS-a	germacrene D (3)
9	***DcTPS34***	Not annotated	TPS-a	
9	***DcTPS65***	Not annotated	TPS-a	
9	***DcTPS20***	Not annotated	TPS-b	
9	***DcTPS39***	DCAR_031040	TPS-a	

(Table 4). No functional characterization of isolated TPS genes *in planta* has been published yet but would be highly desirable in order to analyze TPS genes putatively involved in plant pathogen defense mechanisms. Overexpression of isolated TPS genes in carrot hairy root cultures in vitro might be a possible way to produce large amounts of pharmaceutically relevant terpenes in future.

Compared with the knowledge about carrot terpene profiling, the TPS genes involved, and their putative functions, less was known about the inheritance of carrot terpenes until the studies of Keilwagen et al. (2017) and Le Clerc et al. (2019). However, for targeted breeding, for example, to enhance specific aroma notes or, vice versa, to avoid unwanted terpene profiles leading to bitter taste, it will be necessary to define the most important TPS and other genes from the MEP and MVA pathways involved in the production of major terpenes in carrot roots. Using a combinatorial approach of terpene metabolite profiling, SNP analysis through

genotyping-by-sequencing (GBS), and a subsequent genome-wide association study (GWAS) in a panel of 85 carrot cultivars, Keilwagen et al. (2017) detected 30 QTLs for 15 terpenoid volatiles. Genetic association analysis identified 21 significant QTLs in roots, and 9 QTLs were detected in leaves. Root QTLs were mostly detected for the monoterpenes sabinene, ocimene, β-pinene, borneol and bornyl acetate, and most QTLs were located on carrot chromosomes 4, 5, 7, and 9. Three genomic regions were detected after GWAS, which contained terpene QTLs either associated with a single TPS candidate gene (*DcTPS03* for bornyl acetate and *DcTPS29 for* γ-terpinene) or with the cluster of five mono-TPS genes with high sequence homology on chromosome 4 (*DcTPS04*, *DcTPS26*, *DcTPS27*, *DcTPS54*, and *DcTPS55*) which were associated with QTLs for sabinene and terpinen-4-ol. Analysis of this TPS cluster by QTL mapping in a biparental carrot F_2 population confirmed the results found by GWAS for sabinene and terpinen-4-ol and identified additional QTLs for the monoterpenes α-thujene, α-terpinene, γ-terpinene, and 4-carene in the same genomic region of chromosome 4 (Reichardt et al. 2020). Another previous QTL study, which focused on bitterness trait variation and resistance to *Alternaria dauci* in a segregating carrot F_2 population, identified 71 QTLs for 25 terpenes (Le Clerc et al. 2019). These QTLs were placed mainly on chromosomes 3, 4, and 9, and eight terpenes were the same as in the work of Keilwagen et al. (2017). However, only the QTLs for sabinene on chromosome 4 appeared to be the same in both studies indicating the importance of this genomic region for monoterpene metabolism.

5.5 Implications for Breeding

Marker-assisted breeding (MAS) based on molecular markers linked to QTLs for volatile terpenes might be utilized in carrot breeding programs. Determination of the functional allelic diversity of TPS genes present in *Daucus* germplasm can help to select putative crossing parents. The molecular KASP marker developed by Reichardt et al. (2020) is a first example for such approach (Fig. 9). This allele-specific marker developed from the sequence of the gene *DcTPS54* can discriminate carrot F_2 individuals as well as cultivar genotypes with high or low sabinene content (Reichardt et al. 2020). Sabinene is likely one of the predominant and most important terpenes in carrots, since this substance may be involved in carrot flavor and taste, but was also described to be responsible for bitterness and fungal resistance (Le Clerc et al. 2019). Such correlations between wanted (aroma and resistance) and unwanted (bitterness) traits might lead to conflicts when molecular breeding would ignore such relationships. Therefore, for molecular breeding of high-quality carrot cultivars fulfilling all demands of growers and consumers, specific molecular markers for single terpenoid volatiles would be probably not helpful without a concomitant sensorial perception analysis. Nonetheless, carrot appears as a very promising target crop for breeding programs aiming at the development of specific chemotypes possessing enhanced levels of economically important terpenes for industrial or pharmaceutical exploitations. Special new carrot varieties containing

large amounts of bioactive terpenes but, if necessary, with compromised taste may also be considered as nutraceuticals.

6 Carotenoids

6.1 Introduction

Carotenoids comprise a family of 40-carbon isoprenoids that includes non-oxygenated carotenes as well as oxygenated xanthophylls, among other compounds. They are found in a variety of organisms, including plants, animals, and microbes. Plant carotenoids aid in the absorption of light during photosynthesis and provide protection against photooxidative stress. As accessory pigments, they are essential to development and growth (Lu and Li 2008). In addition, plant growth, flavor, and mycorrhizal connections need carotenoid pathway byproducts (Della Penna and Pogson 2006; Cazzonelli and Pogson 2010). Moreover, carotenoids enhance the plants' non-photosynthetic parts by attracting animals and insects that help pollinate and disperse seeds. It is essential for animals, including humans, to get carotenoids from their diets since with very few exceptions – mammals cannot synthesize carotenoids on their own. These pigments are present in a diverse range of food plants, including carrots, and are the primary precursors of vitamin A, required for immune system function, sight, maturation, and reproduction. In addition, specific carotenoids, such as lycopene and lutein, have been found to reduce the risk of some types of cancers, osteoarthritis, neurodegenerative disorders, and cardiovascular disease (Tanumihardjo 2012).

When it comes to the number and composition of carotenoids in a plant, there is a wide range of variation among tissues and organs within a plant, which also vary during the plant's growth, and across different accessions within a species. Since carotenoids are attractive to animals, their bright colors may have attracted humans' interest as they were domesticated. Multiple regulatory mechanisms are involved in controlling metabolite transport across pathways and cells involved in carotenoid accumulation via biosynthesis (Sun et al. 2018), degradation (Yuan et al. 2015), and sequestration, many of which are characterized by alterations in photomorphogenesis and plastid growth in certain cases (Lu and Li 2008). The recent advances in carrot genomics provided opportunities to expand knowledge about the genetic mechanism controlling carotenoid accumulation in carrot.

Carotenoids are present in very small quantities in the unpigmented roots of wild carrots (*Daucus carota* subsp. *carota*), also called Queen Anne's Lace (QAL), and in white-rooted cultivated carrots, whereas they are present – in varying concentrations – in the majority of the cultivated colored carrots. The primary carotenoid pigments in carrot roots are yellow, orange, and red, and the main compounds that account for such colors are lutein, α- and β-carotene, and lycopene, respectively (Arscott and Tanumihardjo 2010). Yellow carrots predominantly accumulate lutein, zeaxanthin, and α- and β-carotene in small levels (Alasalvar et al. 2001; Arscott and Tanumihardjo 2010), whereas orange carrots typically accumulate large quantities of α- and

β-carotene, and trace amounts of phytoene, lutein, and ζ-carotene. Red carrots accumulate primarily lycopene and often include small amount of α- and β-carotene, as well as lutein (Arscott and Tanumihardjo 2010). Carrot is among the main dietary sources of PACs in many regions of the world, with particular relevance in the USA, where it ranks first in the relative contribution of PACs to the general population's diet. When compared to other dietary sources of PCAs, orange carrots are distinct in that α-carotene can account for a much higher percentage of the total carotenoids content, representing 13–40% of total carotenoids, with higher percentages observed in carrot roots with higher total carotenoid content (Simon and Wolff 1987; Santos and Simon 2006). According to a previous study, carrots can provide 67% of the α-carotene required by the average American diet.

The total carotenoid concentration of dark orange cultivars can exceed 500 ppm on a fresh weight (fw) basis. According to Simon et al. (1989), dark orange cultivars may have a total carotenoid content as high as 500 ppm of fw, and this potential to accumulate carotenoids is associated with the formation of carotenoid structures and crystals in root chromoplasts (Maass et al. 2009; Sun et al. 2018).

Interestingly, there was no report of orange carrots until the seventeenth century. Indeed, when it was first domesticated, about 1100 years ago in Central Asia, carrots had purple or yellow roots (reviewed by Simon 2000). The first orange-rooted carrots were reported in Southern Europe in the sixteenth century and the first red carrot was reported in Asia in the seventeenth century. Since their selection, orange-rooted carrots became – and still are – the predominant market type worldwide. However, red, yellow, and purple carrots still have a relatively large market in some parts of Asia.

6.2 Role of Carrot Carotenoids in Human Nutrition

Orange carrots are rich in carotenes, which predominantly include α- and β-carotene, as well as smaller and varying amounts of γ- and ζ-carotenes, β-zeacarotene, and lycopene, with total carotene content varying from 63 to 548 ppm across different orange-rooted cultivars (Simon and Wolff 1987; Simon 2000). One of the most important functions of carotenoids in the human diet is the role of α-carotene, β-carotene, and cryptoxanthin as precursors of vitamin A. α-carotene and β-carotene account for 13–40% and 45–50% of the total PACs present in orange carrots (Simon and Wolff 1987; Simon 2000). After ingestion, PACs can be cleaved in the intestine or liver, yielding retinaldehyde (retinal) which can then be reduced to retinol, vitamin A (reviewed by Krinsky 1998, and Semba 1998). Vitamin A plays an essential role in human vision and immune function, including apoptosis, keratinization, and B-lymphocyte function (Semba 1998). Vitamin A deficiency can cause a variety of symptoms such as night blindness, xerophthalmia, and increased susceptibility to infections and cancer (Semba 1998). In addition to its ability to produce vitamin A, β-carotene also possesses strong antioxidant activity in humans (Krinsky 1998).

Red-rooted carrots are particularly rich in lycopene, another carotenoid frequently found in the human diet. Lycopene is not a PAC. However, because of the high

number of conjugated double bonds, lycopene has higher antioxidant activity than other carotenoids with fewer double bonds. Several studies have reported an inverse relationship between serum levels of lycopene and risk for several types of cancer, especially prostate cancer, suggesting that lycopene possess anticancer effects (reviewed by Rao and Agarwal 2000). Red carrots are widely consumed in some regions of the world, such as India, where they are preferred over orange carrots.

Yellow carrots are rich in xanthophylls, predominantly lutein and in less amount zeaxanthin. Compared to orange-rooted forms, yellow carrots generally have lower concentration of total carotenoids and β-carotene (Sun et al. 2009). The xanthophyll pigments present in yellow carrots provide protection against cataracts and age-related macular degeneration (Tanumihardjo 2012). These pigments accumulate in both the lens and the macula of the human eye, where they are thought to play a role quenching reactive oxygen species generated by high exposure to ultraviolet light. In addition, it has been indicated that increased intake of lutein may also prevent early atherosclerosis.

White carrots have trace amounts of carotenoids, if any. Because of their extremely low carotenoid content, they lack the carotenoid-related health enhancing properties of the other root color phenotypes. However, the do provide fiber and other nutrients found as well in other colored carrots.

6.3 Biosynthesis of Carrot Carotenoids

Carotenoids are produced in two separate pathways in the plastids and the cytosol. In plastids, the 2-C-methyl D-erythritol-4-phosphorate (MEP) pathway converts pyruvate and glyceraldehyde 3-phosphate to Isopentenyl diphosphate (IPP), and in the cytosol, the mevalonic acid (MVA) pathway converts acetyl-CoA to IPP and geranylgeranyl diphosphate (GGPP). GGPPs are then transformed into phytoene, which is the first step in the carotenoid pathway. Figure 10 presents the MEP and MVA pathways with the respective carrot genes, loci, and enzymes. It is mostly through the MEP route that the majority of carotenoid precursors are made (reviewed by Rodriguez-Concepcion and Stange 2013). There are 44 isoprenoid and 24 carotenoid biosynthesis genes in carrot. Each gene contains various paralogues involved in a variety of modifications, indicating that distinct paralogues evolved specialized functions in different plant taxa, tissue types, developmental stages, environmental conditions, and/or pathways (Rodriguez-Concepcion and Stange 2013; Iorizzo et al. 2016; Simpson et al. 2016). Most of the carotenoid biosynthetic genes are functional in all of the color phenotypes of carrot storage roots, including white-rooted cultivated (Wang et al. 2014; Perrin et al. 2016, 2017a) and wild carrots (Just et al. 2007; Clotault et al. 2008; Bowman et al. 2014; Ma et al. 2017). This is indicative that pathway metabolites serve as precursors for key substances needed for general plant growth and development, such as the hormones abscisic acid and strigolactones.

In carrots with orange, yellow, and red roots, carotenoid concentration increase during root development (Clotault et al. 2008; Fuentes et al. 2012; Perrin et al. 2016; Wang et al. 2014). While the expression of genes increase during root development, it

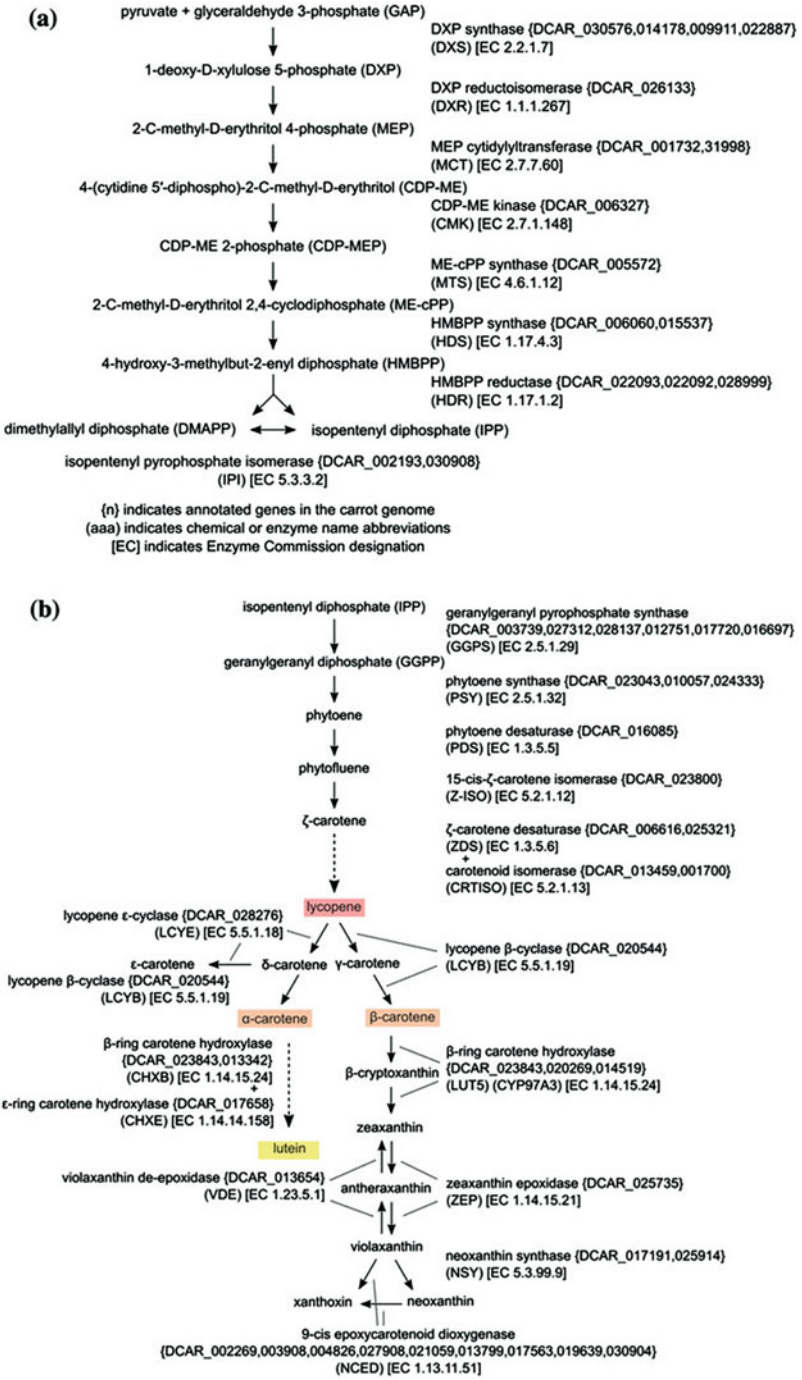


Fig. 10 MEP (2-C-methyl D-erythritol-4-phosphorate) (**a**) and carotenoid (**b**) pathways. Enzyme names, carrot locus tags (in curly brackets), abbreviations (in parentheses), and Enzyme Commission numbers (in square brackets) are included (Reproduced from Simon et al. 2019)

has often been observed that such increase in gene expression is not proportional to – i.e., it is several folds less than – the increase in pigment content (Clotault et al. 2008; Stange et al. 2008; Fuentes et al. 2012; Bowman et al. 2014; Wang et al. 2014; Ma et al. 2017). Greater expression levels of *Lycopene ß-cyclase 1* (*DcLcyb1*) were found in leaves and roots of mature plants than in those organs in immature plants (Moreno et al. 2013). *DcLcyb1* overexpression in transgenic carrots led to increased carotenoid content and increased expression of other carotenoid genes – namely *DcPsy1, DcPsy2, and DcLcyb2* – in leaves and roots. These results suggest that *DcLcyb1* does not possess an organ specific function and modulate carotenoid gene expression and accumulation in carrot leaves and roots.

Transgenic approach has been used by Arango et al. (2014) to study the role of carotene hydroxylase in the metabolism and accumulation of carotenoids in carrot root and leaves. When *CYP97A3* was overexpressed in transgenic orange carrots, the leaves had levels of α-carotene that were about the same as those of untransformed wild carrots. Furthermore, levels of carotenoids were significantly reduced in orange transformed carrots, as was *PSY* protein expression, despite the fact that *PSY* was not expressed in these carrots. This indicated that there is a mechanism that regulates carotenoid metabolism. Indeed, exposure of storage roots to light has a significant impact on the amount of carrot carotenoids and gene expression (Stange et al. 2008; Fuentes et al. 2012).

As with the carotenoid content and the levels of most of the genes involved in the process, the carrot root's shape altered when it was exposed to light in the research. As a consequence of inadequate light and temperature conditions, the carotenoid content of both roots and leaves is decreased (Perrin et al. 2016). The transcriptional control of all of the carotenoid genes in the leaves, as well as the zeaxanthin epoxidase (*ZEP*) and phytoene desaturase (*PDS*) genes in the roots, was held responsible for the observed reduction in AOC. Changes in carotenoid transcript levels, on the other hand, could not explain changes in carotenoid content in contexts with both *Alternaria dauci* infection and water scarcity (Perrin et al. 2017b). This shows that there are other regulatory mechanisms that are not involved in the pathway's functioning.

Due to the wide range of carotenoid content and color intensity found in different carrot genetic stocks and cultivars, many studies have looked at how the expression of carotenoid genes varies in carrots of different colors (Clotault et al. 2008). The enormous and diversified changes in carotenoid composition throughout a wide range of storage root colors are not matched by qualitative differences in gene expression in the pathway, although there are certain similarities in gene expression that are paralleled by carotenoid accumulation patterns (Bowman et al. 2014). White and orange carrots have different levels of two to four times more *PSY1* and *PSY2* genes than white carrots. This is what happened in studies that compared the levels of these two genes in white and orange carrots (Wang et al. 2014). Clotault et al. (2008) reported that yellow carrots have a higher concentration of genes that produce lycopene and ζ-carotene desaturase than orange or white carrots. According to the findings of Ma et al. (2017), yellow cultivars exhibited more genes that produce xanthophyll than orange cultivars. During the period of plant growth in a controlled environment, the expression of carotenoid genes modulates in a way that is comparable to the method in which carotenoid accumulation changes in phloem tissue.

Among genes that encode enzymes are involved in the MEP and MVA pathways, 1-deoxy-d-xylulose-5-phosphate (DXP) synthase 1 (*DXS1*) was the only one to be upregulated in a way that is consistent with the amount of carotenoid present in carrot root (Iorizzo et al. 2016). *DXS* has been identified to be a regulator of isoprenoid biosynthesis in a number of Arabidopsis experiments. According to Simpson et al. (2016), *DXS* is the rate-limiting enzyme in the synthesis of carotenoid pigments in transgenic carrots. In addition, the study found that *PSY* transcript was upregulated in the *DXS* regulatory cascade, which confirms *PSY*'s important role in carotenoid metabolism (Lu and Li 2008; Yuan et al. 2015; Sun et al. 2018). The results that the *PSY* transcript was upregulated in assessments of orange and white carrot roots are likewise consistent with these findings (Bowman et al. 2014). Many studies have also looked at how carotenoid gene sequences change over a wide range of root colors. Carotenoid genes have different nucleotides in carrots from different parts of the world, suggesting that root color was highly relevant in the domestication of carrots. Variation for gene sequence and expression levels were found for seven carotenoid alleles in 46 carrots with different root colors from across the world (Clotault et al. 2012). They found that there was a lot of variation in how much selection was put on genes like *PDS* and *IPP* isomerase (*IPI*), which are at the beginning of the process. The carotene isomerase (*CRTISO*), *LCYB1*, and *LCYE* genes, which are closer to lycopene in the pathway, were also found to be selected for. The sequence variation for *LCYB1* was different between color groups, suggesting that it was selected during domestication. However, there was not a pattern of sequence variation that pointed to candidate genes in the carotenoid pathway that could account for color variation.

Jourdan et al. (2015) examined 17 carotenoid genes in 67 carrot cultivars from across the globe and concluded that there were linkages between α-carotene content and the plastid terminal oxidase (*PTOX*) and *CRTISO* polymorphisms, as well as links between total carotenoid content and the *ZEP*, *PDS*, and *CRTISO* polymorphisms. Genetic linkage between individuals who have the *Y* and Y_2 genes on the same chromosome as *ZEP* and *PDS* may have occurred, which may have contributed them in their connection with one another (Just et al. 2007).

6.4 Carrot Carotenoids Genetics and Genomics

In the late 1960s and early 1980s, Gabelman and his students discovered multiple genes that influence carrot root color. Segregation analysis in populations derived from intercrosses of carrots with different root color revealed that a single dominant gene conditions white root color over yellow, whereas F_2, F_3, and backcross populations from white × orange intercrosses exhibited white and orange-rooted plants consistent with a 2–3 gene patterns of inheritance (Imam and Gabelman 1968; Laferriere and Gabelman 1968). The yellow root color was shown to be more strongly influenced by a monogenic inheritance pattern than the orange root color in this study. One or two additional dominant genes, Y_1 and Y_2, were identified by Kust (1970) as governing white coloration over yellow, as reported by Laferriere and

Gabelman (1968). *O* and *IO*, two genes that enhance orange phloem color, were also discovered by Kust (1970), and further characterized by Buishand and Gabelman (1979), who analyzed segregation of orange color in phloem and xylem tissues in derivatives of orange × white crosses. When the dominant alleles were present in the cross, the carotenoid contents in the offspring of "white × yellow" and "yellow × orange" crosses decreased.

Additionally, Umiel and Gabelman (1972) determined that orange root color predominated over "red white × orange" and "white × yellow" crosses. Two simply inherited loci, *A* and *L*, that influence the accumulation of lycopene and α-carotene were identified by segregation analysis in F_2 and backcross populations derived from intercrosses between red and orange rooted carrots (Umiel and Gabelman 1972). Buishand and Gabelman (1980) found three key genes segregating in progenies from red × yellow crosses: Y_2, which inhibits the synthesis of carotene; *L*, which promotes the synthesis of lycopene; and A_1, which has a phenotypic impact comparable to either *O* or *IO* described by Kust (1970).

Gabelman and colleagues uncovered a number of critical genetic elements influencing root color. Because there were no intercrosses to test for allelism or chromosomal complementation, the only way to distinguish between alleles was to look at the phenotypic traits alone. On the basis of accounts of their characteristics, researchers have emphasized on the *L*, *O*, and *IO* genes.

As previously reported by Arango et al. (2014), the activity of the carotene hydroxylase gene *CYP97A3* in orange carrot roots is reduced, resulting in a higher concentration of α-carotene and higher ratio of α-carotene/β-carotene. This indicates that the "A_1" phenotypic locus described earlier by Buishand and Gabelman (1980) and reported to be essential for α-carotene synthesis in orange carrots, has been identified. A frameshift mutation in the *CYP97A3* gene, which is found only in orange carrot roots, was reported as the causal mutation leading to reduced expression of *CYP97A3* and – thereby – increased accumulation of α-carotene (Arango et al. 2014). This reveals the genetic and molecular foundation for orange carrots' high α-carotene content.

Goldman and Breitbach (1996) described an orange carrot mutant, characterized by "reduce pigment" in the root, caused by the recessive gene "*rp*." This gene reduces the concentration of α- and β-carotene in storage roots by more than 90%, while raising the concentration of phytoene (Koch and Goldman 2005). A distinguishing green color is developed by the sixth leaf in the development of *rprp* plants, which are initially chlorotic and even white. The *rp* mutation results in a decrease in plant vigor, which is rare among genes that regulate carotenoid color in carrots. The carotenoid pathway is thought to be suppressed in the *rp* mutant (Goldman and Breitbach 1996); however, this hypothesis has yet to be tested. Because of its uniqueness, further characterization of *rp* might provide new insights into the metabolism of carrot carotenoids.

In three orange × yellow crosses, Simon (1996) used carotenoid mapping populations to examine the relationship between Y_2 and genes that govern sugar and anthocyanin accumulation. He found that Y_2 had a monogenic inheritance pattern. Because the observed phenotype most closely correlated with Buishand and

Gabelman's (1979) phenotypic descriptions of this trait, the gene was designated as Y_2. An AFLP fragment genetically linked to Y_2 was converted into a PCR-based marker to help in the phenotyping of carotenoid accumulation genes in breeding programs (Bradeen and Simon 1998). AFLPs were initially used to map the Y_2 gene, which conditions yellow versus orange root color and carotene concentration (Vivek and Simon 1999). Santos and Simon (2002) employed AFLP markers to map QTLs for carotenoids concentration (assessed by HPLC analysis) in two mapping populations. Significant QTLs were identified for several carotenoid products, including phytoene, lycopene, β-carotene, and α-carotene. Eleven QTLs for concentration of α- and β-carotene were detected and they were associated with orange root color.

Carotenoid accumulation genes were formerly thought to be metabolic pathway structural genes, like in other crops. It is possible that the genes conditioning carotenoid-related coloring of carrot roots are not structural biosynthetic genes but rather control carotenoid accumulation (Santos et al. 2005). A single major QTL for β-carotene, total carotene, and lycopene content was identified in the F_2 population of P50006 and HCM A.C. using sequence related amplified polymorphism (SRAP) markers, and this QTL accounted for ~12.8–14.6% of the total phenotypic variation for the three pigment traits. The genetic variability of these three QTLs was attributed to additive genetic variance rather than additive genetic variation. Furthermore, a pair of epistatic QTLs for β-carotene and lycopene accumulation was shown to account for 15.1% and 6.5% of the total phenotypic variance, respectively. The SRAP markers tightly linked to these QTLs might be employed in carrot breeding to select for high carotenoids and lycopene content via QTLs pyramiding or selection for high carotenoids and lycopene content through selection.

Santos and Simon (2006) also estimated broad-sense heritability values for individual carotenoids and cumulative carotenoid concentration in the two populations. A segregating pattern was revealed in the progeny of a hybrid between a wild, white-rooted carrot, known as QAL, and a dark orange cultivated carrot, known as B493, in which the *Y* and Y_2 genes were segregating, as well as quantitative loci that contributed to variation in total carotenoid concentration.

In the same study of Santos and Simon (2006), progenies from another cross between the orange-rooted Brasilia and the dark orange HCM were examined. Although the genotypes of the progenies were all recessive for the *Y* and Y_2 loci (yyy_2y_2), the progenies were quantitatively segregating for carotene content, in approximately fivefold difference between the two parents. The total heritability of carotenoid content in the 'B493 × QAL' population ranged from 0.89 to 0.98, while it ranged from 0.38 to 0.45 in the 'Brasilia × HCM' background, demonstrating that the *Y* and Y_2 genes had a significant influence on descendants in the B494 × QAL population. Carotenoid concentrations in the same orange carrot genetic stocks may vary up to twofolds in different environments, according to Simon and Wolff (1987) and Perrin et al. (2016), explaining why the 'Brasilia × HCM' population had lower heritability than the other HCM populations. It was hypothesized that the genes encoding the carotenoid biosynthesis enzymes could be linked to the carotenoid color genes, hence Just et al. (2007, 2009) identified and mapped 22 probable carotenoid enzyme genes in carrot. The population used in this research was generated from a cross between a wild

white-rooted carrot known as QAL and a cultivated carrot known as B493, which has a dark orange root. This population segregated for orange, yellow, and white root colors, with ratios consistent with a two-gene model. Bradeen and Simon (1998) genetically mapped the Y_2 gene on linkage group 5 and *Y* on linkage group 2. *ZEP* and *ZDS* were found to be positionally associated with the Y_1 and Y_2 genes, respectively. The Y_1 gene was shown to be associated with ε-ring carotene hydroxylase (*CHXE*), 9-cis-epoxycarotenoid dioxygenase 2 (*NCED2*), and phytoene desaturase (PDS). Due to the lack of close relationships, they were designated positional candidates (Just et al. 2009). After sequencing the carrot genome (Iorizzo et al. 2016), a *Y* gene candidate was found. Fine mapping in this study revealed a 75-kb region of chromosome 5 responsible for both the orange and pale orange (YYy_2y_2) root color segregating as a monogenic trait in the B493 × QAL cross population, as well as the yellow (yyY_2Y_2) and white (YYY_2Y_2) root color segregating in an unrelated population. In this study, the light orange phenotype (yyy_2y_2) has not previously been connected to aYYy_2y_2 genotype, since earlier phenotyping had not been able to differentiate it from orange (yyy_2y_2). The identification of a potential *Y* locus gene was made using differential expression analysis and nucleotide polymorphisms (*DCAR_032551*) (Iorizzo et al. 2016). It was observed that this gene co-expressed with *DXS1* and *LCYE*, two genes engaged in the isoprenoid pathway, in non-photosynthetic root tissue, along with a slew of other genes involved in photosynthetic induced changes and functioning, plastid biosynthesis, and chlorophyll metabolism. Arabidopsis has a pseudo-etiolated morphology and interacts with genes that control light response and photomorphogenesis, including the *Y* gene potential homolog pseudo-etiolation in light (Ichikawa et al. 2006). *DXS1* is involved in the production of carotenoid precursors in photosynthetic metabolism and is activated by light (Stange et al. 2008; Fuentes et al. 2012). *DXS1* overexpression and carotenoid synthesis may be facilitated by a recessive allele (*yy*) in etiolated roots, according to a study by Iorizzo et al. (2016). Orange carrot roots exposed to light have been shown to undergo morphological changes that might be explained by this hypothesis, which is currently being tested out (Stange Klein and Rodriguez-Concepcion 2015).

The Y_2 and *Y* genes were shown to be segregating in the 'B493 × QAL'-derived population (Just et al. 2009). By combining molecular mapping with transcriptome analysis, it was feasible to determine the location of the gene Y_2 in a population that was homozygous recessive for *Y* (*yy*) but segregated for root color associated with the Y_2 region (Ellison et al. 2017). Many carotenoid-related genes were found to be differentially expressed between orange and yellow roots at 40 and 80 days following planting. A total of 6 genes were discovered in the 650-kb region, including *PSY1, PSY3*, geranylgeranyl diphosphate synthase 1 (*GPPS1*), neoxanthin synthase 1 (*NSY1*), carotenoid cleavage dioxygenase 1 (*CCD1*), *LUTEIN DEFICIENT 5 (LUT5)*, and two cytochrome genes. In this study, it was observed that DXP reductoisomerase (DXR) was the only MEP or carotenoid pathway gene expressed present in that region, despite the fact that it did not express differently in orange versus yellow carrots. Among the 17 genes that changed between 40 and 80 days were detected in the fine-mapped region, only 4 were found in the fine-mapped region. A recessive phenotype was anticipated for just one of the four, the "protein

dehydration-induced 19" homolog 5 *(Di19)* (DCAR 026175) gene, which showed lower expression in orange roots than yellow roots. *AtDi19-7*, a member of the Arabidopsis *Di19* gene family, has been associated with light signaling responses, and other abiotic stimuli, including abscisic acid (Milla et al. 2006). Variations in the expression of *Di19* during photomorphogenesis may have an impact on the synchronized synthesis of chlorophyll and carotenoids throughout this process. In order to confirm the Y_2 candidate, more testing is needed since the 650-kb region reported by Ellison et al. (2017) contains several candidate genes.

Another gene that impacts carotenoid accumulation in carrot roots and includes the whole spectrum of carotenoid colors was identified in a recent association research that included data from 154 wild and 520 cultivated carrots from geographically diverse worldwide growing regions (Ellison et al. 2018). This carrot collection was analyzed using genotyping by sequencing (GBS) to look for evidence of domestication. The study discovered a connection between high levels of carotene and the existence of the *Or* gene in a 143-kb genomic region of chromosome 3, which lacked any MEP or carotenoid genes. *Or* is essential for chromoplast formation, which acts as a sink for carotenoids in several plants species, including Arabidopsis, cauliflower, and sweet potato (Lu and Li 2008; Yuan et al. 2015; Sun et al. 2018), and a similar role for *Or* in carrot root has been hypothesized (Lu and Li 2008; Yuan et al. 2015; Sun et al. 2018). Cultivated carrots from Central Asia, which is the principal source of carrot diversification (Iorizzo et al. 2013), had higher expression levels of the *Or* gene than cultivated carrots from Europe. While the wild-type allele for *Or* (*Orw*) causes yellow root color, recessive genotypes present light orange root, and the storage roots of $OrcOrcyyy_2y_2$ plants are yellow in color, while the wild-type allele for *Or* (*Orw*) causes yellow color. *Orc* allele polymorphism may have had a role in carrot domestication in Central Asia prior to the *Or* gene being fixed for use in European carrots. Scientists are currently investigating the relationship between *Or* allele-specific gene expression and phenotypes.

6.5 Genetic Engineering for Enhancing Carotenoids Levels in Carrot

Increased carotenoid production and accumulation were achieved by introducing the phytoene synthase (*PSY*) gene from *Erwinia herbicola* into the carrot plant and combining it to a plastid transit peptide, which targets the enzyme that targets chromoplasts (Hauptmann et al. 1997). Its expression enhanced phytoene biosynthesis in orange carrot root, which resulted in ~twofold increase in β-carotene concentration. When the *crt*B gene under the control of a yam root-specific promoter was genetically engineered into wild carrot, the color of the transformed roots shifted from white to brilliant yellow, indicating that the gene was successfully expressed. A high concentration of carotene intermediates, such as phytoene, phytofluene, and ζ-carotene and lycopene, which accompanied β-carotene in significantly lower proportions than in normal orange root, were responsible for the different root colors, whereas α-carotene was not detected in this sample. In this way, upregulation of *PSY* in resulted in higher

carotenoid concentrations in carrot roots and enhanced carotenoid storage in chromoplasts in the crystalline form, similar to what occurs in orange carrots, despite the fact that the carotenoids composition was different (Maass et al. 2009).

The carrot homologue of the carotene hydroxylase gene from Arabidopsis*CYP97A3* was found to be expressed in orange carrot roots as well as in other plants. However, while the Arabidopsis gene (*AtCYP97A3*) is fully functional, the carrot *DcCYP97A3* present in orange-rooted cultivars is only partially functional or nonfunctional, due to an 8-nucleotide frameshift insertion that generated a premature stop codon, resulting in the production of a truncated hydroxylase protein (Arango et al. 2014). As a result, the conversion of α-carotene to lutein is inhibited, therefore α-carotene content and the α/β carotene ratio are increased (Fig. 1). The overexpression of *AtCYP97A3* in transgenic orange carrots resulted in a reduction in the quantity of α-carotene present in the carotenoid composition. It was also noticed that carotene hydroxylase overexpression was associated with a reduction in total carotenoids, which was due to a lower *PSY* protein level, despite no changes observed in *PSY* gene expression. It was hypothesized that carotene hydroxylase overexpression may negatively affect *PSY* protein translation (Arango et al. 2014).

Similarly, the significance of precursors in carotenoid biosynthesis was investigated by introducing – via transgenics – the Arabidopsis deoxyxylulose 5-phosphate synthase (*DXS*) and reductoisomerase (*DXR*) genes from the MEP pathway in carrot. The overexpression of *DXS* increased *PSY* transcript levels and, as result, increased carotenoid quantities by an average of twofold, but the overexpression of *DXR* had no significant impact (Simpson et al. 2016).

The effect of the algal *Haematococcus pluvialis* β-carotene ketolase (*bkt*) on carrot carotenoids production was investigated. In the presence of this enzyme, β-carotene is converted to ketocarotenoids such as canthaxanthin and astaxanthin, which have great antioxidant effects and are useful nutraceuticals in the human diet, as well as feed additives for pale fish, salmon, and trout cultures. It was necessary to fuse the bkt gene to the ribulose bisphosphate carboxylase-oxygenase (RuBisCO) signal peptide in order to assure that the enzyme would function in plastids. Leaf and root levels of carrot β-carotene hydroxylases were upregulated in comparison to other tissues. The accumulation in carrot roots of ketocarotenoids, primarily astaxanthin, adonirubin, and canthaxanthin, up to concentrations of 2400μg/g dry weight, was seen in conjunction with a decline in carotenes (Jayaraj and Punja 2008). The growth of plants having high concentrations of ketocarotenoids was improved when they were subjected to intense UV-B irradiation, and the leaves were less damaged when H_2O_2 or methyl viologen stress was administered to them. As a result of the significant antioxidant and free radical scavenging activity of ketocarotenoids, it was inferred that the cells were protected from oxidative stress (Jayaraj and Punja 2007).

6.6 Perspectives on Carrot Carotenoids

Carrot roots may store a variety of carotenoids via divergent pathways. The study of carotenoid production and accumulation in carrot storage roots employs a

multidisciplinary approach that includes genetics, gene expression, and metabolic interactions. There are three primary root colors associated with the accumulation of carotenoids in this organ: yellow, orange, and red. Each color represents a very simple set of metabolic activities that are linked to variations in genetic differences for the genes involved, expression levels, and carotenoid accumulation throughout periods of development. However, when looking at the link between root color and carotenoid content, it has been difficult to figure out how the carotenoid pathway may account for the variations observed in carotenoid composition in orange, yellow, and red carrots. The color of carotenoid pigments is affected by variations in the way they are synthesized, as well as by changes in photomorphogenesis and plastid development. We know relatively little about the regulatory genes involved in the non-carotenoid biosynthesis pathway, discovered by means of carrot genome sequencing and other multi-omics approaches, and we hope to learn more in the future.

References

Aarts MG, Keijzer CJ, Stiekema WJ, Pereira A (1995) Molecular characterization of the *CER1* gene of Arabidopsis involved in epicuticular wax biosynthesis and pollen fertility. Plant Cell 7:2115–2127

Aćimović M, Stanković J, Cvetković M, Ignjatov M, Nikolić L (2016) Chemical characterization of essential oil from seeds of wild and cultivated carrots from Serbia. Bot Serb 40:44–60

Alasalvar C, Grigor JM, Zhang D, Quantick PC, Shahidi F (2001) Comparison of volatiles, phenolics, sugars, antioxidant vitamins, and sensory quality of different colored carrot varieties. J Agric Food Chem 49:1410–1416

Alegria C, Gonçalves EM, Moldão-Martins M, Cisneros-Zevallos L, Abreu M (2016) Peel removal improves quality without antioxidant loss, through wound-induced phenolic biosynthesis in shredded carrot. Postharvest Biol Technol 120:232–239

Algarra M, Fernandes A, Mateus N, de Freitas V, Esteves da Silva J et al (2014) Anthocyanin profile and antioxidant capacity of black carrots (*Daucus carota* L. ssp. *sativus* var. *atrorubens* Alef.) from Cuevas Bajas, Spain. J Food Compos Anal 33:71–76

Alves-Silva JM, Zuzarte M, Gonçalves MJ, Cavaleiro C, Cruz MT et al (2016) New claims for wild carrot (*Daucus carota* subsp. *carota*) essential oil. Evid Based Complementary Altern Med 2016:9045196

Andersen OM, Jordheim M (2006) The anthocyanins. In: Andersen OM, Markham KR (eds) Flavonoids chemistry, biochemistry and applications. CRC Press/Taylor and Francis, Boca Raton, pp 471–551

Arango J, Jourdan M, Geoffriau E, Beyer P, Welsch R (2014) Carotene hydroxylase activity determines the levels of both alpha-carotene and total carotenoids in orange carrots. Plant Cell 26:2223–2233

Arscott SA, Tanumihardjo SA (2010) Carrots of many colors provide basic nutrition and bioavailable phytochemicals acting as a functional food. Compr Rev Food Sci Food Saf 9:223–239

Bahi A, Al Mansouri S, Al Memari E, Al Ameri M, Nurulain SM et al (2014) β-Caryophyllene, a CB2 receptor agonist produces multiple behavioral changes relevant to anxiety and depression in mice. Physiol Behav 135:119–124

Balasundram N, Sundram K, Samman S (2006) Phenolic compounds in plants and Agri-industrial by-products: antioxidant activity, occurrence, and potential uses. Food Chem 99:191–203

Bannoud F, Da Peña Hamparsomián J, Insani M, Chamorro V, Pazos A, et al. (2018) Assessment of genetic diversity for root anthocyanin composition and phenolic content in purple carrots. In: 2nd International symposium on carrot and other apiaceae, abstracts book. Krakow, Poland, September 19–22, pp 45–46

Bannoud F, Ellison S, Paolinelli M, Horejsi T, Senalik D et al (2019) Dissecting the genetic control of root and leaf tissue-specific anthocyanin pigmentation in carrot (*Daucus carota* L.). Theor Appl Genet 132:2485–2507

Bannoud F, Carvajal S, Ellison S, Senalik D, Gomez Talquenca S et al (2021) Genetic and transcription profile analysis of tissue-specific anthocyanin pigmentation in carrot root phloem. Genes 12:1464–1483

Baranska M, Schulz H, Baranski R, Nothnagel T, Christensen L (2005) In situ simultaneous analysis of polyacetylenes, carotenoids and polysaccharides in carrot roots. J Agric Food Chem 53:6565–6571

Baranski R, Lukasiewicz A (2019) Genetic engineering of carrot. In: Simon P, Iorizzo M, Grzebelus D, Baranski R (eds) The carrot genome. Springer, Cham, pp 149–186

Baranski R, Maksylewicz-Kaul A, Nothnagel T, Cavagnaro PF, Simon PW et al (2012) Genetic diversity of carrot (*Daucus carota* L.) cultivars revealed by analysis of SSR loci. Genet Resour Crop Evol 59:163–170

Bartley GE, Avena-Bustillos RJ, Du W-X, Hidalgo M, Cain B et al (2016) Transcriptional regulation of chlorogenic acid biosynthesis in carrot root slices exposed to UV-B light. Plant Gene 7:1–10

Becerra-Moreno A, Redondo-Gil M, Benavides J, Nair V, Cisneros-Zevallos L et al (2015) Combined effect of water loss and wounding stress on gene activation of metabolic pathways associated with phenolic biosynthesis in carrot. Front Plant Sci 6:837

Behrens CE, Smith KE, Iancu CV, Choe JY, Dean JV (2019) Transport of anthocyanins and other flavonoids by the Arabidopsis ATP-binding cassette transporter *AtABCC2*. Sci Rep 9:437

Bowman MJ, Willis DK, Simon PW (2014) Transcript abundance of phytoene synthase 1 and phytoene synthase 2 is association with natural variation of storage root carotenoid pigmentation in carrot. J Am Soc Hortic Sci 139:63–68

Bradeen JM, Simon PW (1998) Conversion of an AFLP fragment linked to the carrot Y2 locus to a simple, codominant PCR-based marker form. Theor Appl Genet 97:960–967

Buishand JG, Gabelman WH (1979) Investigations on the inheritance of color and carotenoid content in phloem and xylem of carrot roots (*Daucus carota* L.). Euphytica 28:611–632

Buishand JG, Gabelman WH (1980) Studies on the inheritance of root color and carotenoid content in red × yellow and red × white crosses of carrot, *Daucus carota* L. Euphytica 29:241–260

Busta L, Yim WC, LaBrant EW, Wang P, Grimes L et al (2018) Identification of genes encoding enzymes catalyzing the early steps of carrot polyacetylene biosynthesis. Plant Physiol 178: 1507–1521

Cahoon EB, Schnurr JA, Huffman EA, Minto RE (2003) Fungal responsive fatty acid acetylenases occur widely in evolutionarily distant plant families. Plant J 34:671–683

Carnesecchi S, Bras-Gonçalves R, Bradaia A, Zeisel M, Gossé F et al (2004) Geraniol, a component of plant essential oils, modulates DNA synthesis and potentiates 5-fluorouracil efficacy on human colon tumor xenografts. Cancer Lett 215:53–59

Cavagnaro PF (2019) Genetics and genomics of carrot sugars and polyacetylenes. In: Simon P, Iorizzo M, Grzebelus D, Baranski R (eds) The carrot genome. Springer, Cham, pp 295–315

Cavagnaro PF, Iorizzo M (2019) Carrot anthocyanins. Diversity, genetics and genomics. In: Simon P, Iorizzo M, Grzebelus D, Baranski R (eds) The carrot genome. Springer, Cham, pp 261–277

Cavagnaro PF, Iorizzo M, Yildiz M, Senalik D, Parsons J et al (2014) A gene-derived SNP-based high resolution linkage map of carrot including the location of QTLs conditioning root and leaf anthocyanin pigmentation. BMC Genomics 15:1118

Cavagnaro P, Bannoud F, Iorizzo M, Senalik D, Ellison S et al (2019) Carrot anthocyanins: nutrition, diversity and genetics. Acta Hortic 1264:101–106

Cazzonelli CI, Pogson BJ (2010) Source to sink: regulation of carotenoid biosynthesis in plants. Trends Plant Sci 15:266–274

Charron CS, Kurilich AC, Clevidence BA et al (2009) Bioavailability of anthocyanins from purple carrot juice: effects of acylation and plant matrix. J Agric Food Chem 57:1226–1230

Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J 66:212–229

Chen X, Chen H, Yuan JS, Köllner TG, Chen Y et al (2018) The rice terpene synthase gene *OsTPS19* functions as an (S)-limonene synthase in planta, and its overexpression leads to enhanced resistance to the blast fungus *Magnaporthe oryzae*. Plant Biotechnol J 16: 1778–1787

Christensen LP (2011) Aliphatic C(17)-polyacetylenes of the falcarinol type as potential health promoting compounds in food plants of the Apiaceae family. Recent Pat Food Nutr Agric 3:64–77

Christensen LP (2020) Bioactive C17 and C18 acetylenic oxylipins from terrestrial plants as potential lead compounds for anticancer drug development. Molecules 25:2568

Christensen LP, Brandt K (2006) Bioactive polyacetylenes in food plants of the Apiaceae family: occurrence, bioactivity and analysis. J Pharm Biomedical Anal 41:683–693. 19

Christensen LP, Kreutzmann S (2007) Determination of polyacetylenes in carrot roots (*Daucus carota* L.) by high-performance liquid chromatography coupled with diode array detection. J Sep Sci 30:483–490

Clotault J, Peltier D, Berruyer R, Thomas M, Briard M et al (2008) Expression of carotenoid biosynthesis genes during carrot root development. J Exp Bot 59:3563–3573

Clotault J, Peltier D, Soufflet-Freslon V, Briard M, Geoffriau E (2012) Differential selection on carotenoid biosynthesis genes as a function of gene position in the metabolic pathway: a study on the carrot and dicots. PLoS One 7:e38724

Cox-Georgian D, Ramadoss N, Dona C, Basu C (2019) Therapeutic and medicinal uses of terpenes. Med Plants 12:333–359

Crowell PL (1999) Prevention and therapy of cancer by dietary monoterpenes. J Nutr 129: 775S–778S

Curaba J, Bostan H, Cavagnaro PF, Senalik D, Mengist MF et al (2020) Identification of an SCPL gene controlling anthocyanin acylation in carrot (*Daucus carota* L.) root. Front Plant Sci 10:1770

Czepa A, Hofmann T (2004) Quantitative studies and sensory analyses on the influence of cultivar, spatial tissue distribution, and industrial processing on the bitter off-taste of carrots (*Daucus carota* L.) and carrot products. J Agric Food Chem 52:4508–4514

Dawid C, Dunemann F, Schwab W, Nothnagel T, Hofmann T (2015) Bioactive C_{17}-polyacetylenes in carrots (*Daucus carota* L.): current knowledge and future perspectives. J Agric Food Chem 63:9211–9222

de Rosas I, Deis L, Baldo Y, Cavagnaro JB, Cavagnaro PF (2022) High temperature alters anthocyanin concentration and composition in grape berries of malbec, merlot, and pinot noir in a cultivar-dependent manner. Plan Theory 11:926

Deding U, Baatrup G, Christensen LP, Kobaek-Larsen M (2020) Carrot intake and risk of colorectal cancer: a prospective cohort study of 57,053 Danes. Nutrients 12:332

Dees C, Askari M, Garret S (1997) Estrogenic and DNA-damaging activity of red no.3 in human breast cancer cells. Environ Health Perspect 105:625–632

Della Penna D, Pogson BJ (2006) Vitamin synthesis in plants: Tocopherols and carotenoids. Annu Rev Plant Biol 57:711–738

Dunemann F, Böttcher C (2021) Enhanced production of falcarinol-type polyacetylenes in hairy roots of cultivated carrot (*Daucus carota* subsp. *sativus*). Preprint, Research Square. https://doi.org/10.21203/rs.3.rs-653311/v3

Dunemann F, Ulrich D, Böttcher C, Budahn H, Lehnert H et al (2019) Genomics-assisted analysis of natural substances in carrots. Acta Hortic 1264:91–100

Dunemann F, He W, Böttcher C, Reichardt S, Nothnagel T et al (2022) The genetic control of polyacetylenes involved in bitterness of carrots (*Daucus carota* L.): identification of QTLs and candidate genes from the plant fatty acid metabolism. BMC Plant Biol 22:92

El-Houri RB, Mortier J, Murgueitio MS, Wolber G, Christensen LP (2015) Identification of PPARgamma agonists from natural sources using different *in silico* approaches. Planta Med 81:488–494

Ellison S, Senalik D, Bostan H, Iorizzo M, Simon P (2017) Fine mapping, transcriptome analysis, and marker development for Y_2, the gene that conditions beta-carotene accumulation in carrot (*Daucus carota* L). Genes Genomes Genet 7:2665–2675

Ellison SL, Luby CH, Corak K, Coe K, Senalik D et al (2018) Carotenoid presence is associated with the *or* gene in domesticated carrot. Genetics 210:1–12

Falara V, Akhtar TA, Nguyen TTH, Spyropoulou EA, Bleeker PM et al (2011) The tomato terpene synthase gene family. Plant Physiol 157:770–789

Feller A, Machemer K, Braun EL, Grotewold E (2011) Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. Plant J 66:94–116

Flamini G, Cosimi E, Cioni PL, Molfetta I, Braca A (2014) Essential oil composition of *Daucus carota* ssp. *major* (Pastinocello carrot) and nine different commercial varieties of *Daucus carota* ssp. *Sativus* fruits. Chem Biodivers 11:1022–1033

Formica-Oliveira AC, Martínez-Hernández GB, Aguayo E, Gómez PA, Artés F et al (2016) UV-C and hiperoxia abiotic stresses to improve healthiness of carrots: study of combined effects. J Food Sci Technol 53:3465–3476

Fuentes P, Pizarro L, Moreno JC, Handford M, Rodriguez-Concepcion M et al (2012) Light-dependent changes in plastid differentiation influence carotenoid gene expression and accumulation in carrot roots. Plant Mol Biol 79:47–59

Fukuda T, Okazaki K, Shinano T (2013) Aroma characteristic and volatile profiling of carrot varieties and quantitative role of terpenoid compounds for carrot sensory attributes. J Food Sci 78:S1800–S1806

Fukumoto LR, Mazza G (2000) Assesing antioxidant and prooxidant activities of phenolic compounds. J Agric Food Chem 48:3597–3604

Gertsch J, Leonti M, Raduner S, Racz I, Chen JZ et al (2008) Beta-caryophyllene is a dietary cannabinoid. Proc Natl Acad Sci USA 105:9099–9104

Goldman IL, Breitbach DN (1996) Inheritance of a recessive character controlling reduced carotenoid pigmentation in carrot (*Daucus carota* L.). J Hered 87:380–382

Gomez C, Terrier N, Torregrosa L, Vialet S, Alexandre FL et al (2009) Grapevine MATE-type proteins act as vacuolar H+-dependent acylated anthocyanin transporters. Plant Physiol 150:402–415

Güler Z, Karaca F, Yetisir H (2015) Identification of volatile organic compounds (VOCs) in different colour carrot (*Daucus carota* L.) cultivars using static headspace/gas chromatography/mass spectrometry. Cogent Food Agric 1:1117275

Guzmán-Gutiérrez SL, Reyes-Chilpa R, Jaime HB (2014) Medicinal plants for the treatment of "nervios", anxiety, and depression in Mexican traditional medicine. Rev Bras 24:591–608

Habegger R, Schnitzler WH (2000) Aroma compounds in the essential oil of carrots (*Daucus carota* L. ssp. *sativus*). 1. Leaves in comparison with roots. J Appl Bot 74:220–223

Hampel D, Mosandl A, Wüst M (2005) Biosynthesis of mono- and sesquiterpenes in carrot roots and leaves (*Daucus carota* L.): metabolic cross talk of cytosolic mevalonate and plastidial methylerythritol phosphate pathways. Phytochemistry 66:305–311

Hauptmann R, Eschenfeldt WH, English J, Brinkhaus FL (1997) Enhanced carotenoid accumulation in storage organs of genetically engineered plants. US Patent 5:618–988

He J, Giusti MM (2010) Anthocyanins: natural colorants with health-promoting properties. Annu Rev Food Sci Technol 1:163–187

Herrmann KM, Weaver LM (1999) The shikimate pathway. Annu Rev Plant Physiol Plant Mol Biol 50:473–503

Heydenreuter W, Kunold E, Sieber SA (2015) Alkynol natural products target ALDH2 in cancer cells by irreversible binding to the active site. Chem Commun 51:15784–15787

Holton TA, Cornish EC (1995) Genetics and biochemistry of anthocyanin biosynthesis. Plant Cell 7:1071–1083

Ibdah M, Muchlinski A, Yahyaa M, Nawade B, Tholl D (2019) Carrot volatile terpene metabolism: terpene diversity and biosynthetic genes. In: Simon P, Iorizzo M, Grzebelus D, Baranski R (eds) The carrot genome. Springer, Cham, pp 247–260

Ichikawa T, Nakazawa M, Kawashima M, Iizumi H, Kuroda H et al (2006) The FOX hunting system: an alternative gain-of-function gene hunting technique. Plant J 48:974–985

Imam MK, Gabelman WH (1968) Inheritance of carotenoids in carrots, *Daucus carota*, L. Proc Am Soc Hortic Sci 93:419–428

Iorizzo M, Senalik DA, Ellison SL, Grzebelus D, Cavagnaro PF et al (2013) Genetic structure and domestication of carrot (*Daucus carota* L. subsp. *sativus* L.) (Apiaceae). Am J Bot 100:930–938

Iorizzo M, Ellison S, Senalik D, Zeng P, Satapoomin P et al (2016) A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. Nat Genet 48:657–666

Iorizzo M, Cavagnaro PF, Bostan H, Zhao Y, Zhang J et al (2019) A cluster of MYB transcription factors regulate anthocyanin biosynthesis in carrot (*Daucus carota* L.) root and leaf. Front Plant Sci 9:1927

Iorizzo M, Curaba J, Pottorff M, Simon PW, Cavagnaro PF (2020) Carrot anthocyanin genetics: status and perspectives to optimize applications for the food colorant industry. Genes 11:906

Jabrane A, Jannet HB, Harzallah-Skhiri F, Mastouri M, Casanova J et al (2009) Flower and root oils of the Tunisian *Daucus carota* L. ssp. *maritimus* (Apiaceae): integrated analyses by GC, GC/MS, and 13C-NMR spectroscopy, and *in vitro* antibacterial activity. Chem Biodivers 6:881–889

Jasicka-Misiak I, Lipok J, Nowakowska EM, Wieczorek PP, Młynarz P et al (2014) Antifungal activity of the carrot seed oil and its major sesquiterpene compounds. Z Naturforsch 59:791–766

Jayaraj J, Punja ZK (2007) Combined expression of chitinase and lipid transfer protein genes in transgenic carrot plants enhances resistance to foliar fungal pathogens. Plant Cell Rep 26: 1539–1546

Jayaraj J, Punja ZK (2008) Transgenic carrot plants accumulating ketocarotenoids show tolerance to UV and oxidative stresses. Plant Physiol Biochem 46:875–883

Jeon JE, Kim JG, Fischer CR, Mehta N, Dufour-Schroif C et al (2020) A pathogen-responsive gene cluster for highly modified fatty acids in tomato. Cell 180:176–187

Jha NK, Sharma C, Hashiesh HM, Arunachalam S, Meeran MFN et al (2021) β-Caryophyllene, a natural dietary CB2 receptor selective cannabinoid can be a candidate to target the trinity of infection, immunity, and inflammation in COVID-19. Front Pharm 12:590201

Jing P, Bomser JA, Schwartz SJ, He J, Magnuson BA et al (2008) Structure-function relationships of anthocyanins from various anthocyanin-rich extracts on the inhibition of colon cancer cell growth. J Agric Food Chem 56:9391–9398

Jourdan M, Gagne S, Dubois-Laurent C, Maghraoui M, Huet S et al (2015) Carotenoid content and root color of cultivated carrot: a candidate-gene association study using an original broad unstructured population. PLoS One 10:e0116674

Just BJ, Santos CAF, Fonseca MEN, Boiteux LS, Oloizia BB et al (2007) Carotenoid biosynthesis structural genes in carrot (*Daucus carota*): isolation, sequence-characterization, single nucleotide polymorphism (SNP) markers and genome mapping. Theor Appl Genet 114:693–704

Just BJ, Santos CA, Yandell BS, Simon PW (2009) Major QTL for carrot color are positionally associated with carotenoid biosynthetic genes and interact epistatically in a domesticated × wild carrot cross. Theor Appl Genet 119:1155–1169

Kammerer D, Carle R, Schieber A (2004) Quantification of anthocyanins in black carrot extracts (*Daucus carrota* ssp. *sativus* var. *atrorubens* Alef.) and evaluation of their colour properties. Eur Food Res Technol 219:479–486

Keilwagen J, Lehnert H, Berner T, Budahn H, Nothnagel T et al (2017) The terpene synthase gene family of carrot (*Daucus carota* L.): identification of QTLs and candidate genes associated with terpenoid volatile compounds. Front Plant Sci 8:1930

Khalil N, Ashour M, Singab AN, Salama O (2015) Chemical composition and biological activity of the essential oils obtained from yellow and red carrot fruits cultivated in Egypt. IOSR J Pharm Biol Sci 10:13–19

Kjeldsen F, Christensen LP, Edelenbos M (2003) Changes in volatile compounds of carrots (*Daucus carota* L.) during refrigerated and frozen storage. J Agric Food Chem 51:5400–5407

Kobaek-Larsen M, Baatrup G, Notabi MK, El-Houri RB, Pipó-Ollé E et al (2019) Dietary polyacetylenic oxylipins falcarinol and falcarindiol prevent inflammation and colorectal neoplastic transformation: a mechanistic and dose-response study in a rat model. Nutrients 11:2223

Koch T, Goldman IL (2005) Relationship of carotenoids and tocopherols in a sample of carrot root-color accessions and carrot germplasm carrying *Rp* and *rp* alleles. J Agric Food Chem 53: 325–331

Kodama M, Brinch-Pedersen H, Sharma S, Holme IB, Joernsgaard B et al (2018) Identification of transcription factor genes involved in anthocyanin biosynthesis in carrot (*Daucus carota* L.) using RNA-Seq. BMC Genomics 19:811

Koutouan C, Le Clerc V, Baltenweck R, Claudel P, Halter D et al (2018) Link between carrot leaf secondary metabolites and resistance to *Alternaria dauci*. Sci Rep 8:13746

Koutsogeorgopoulou L, Maravelias C, Methenitou G, Koutselinis A (1998) Immunological aspects of the common food colorants, amaranth and tartrazine. Vet Hum Toxicol 40:1–4

Krähmer A, Böttcher C, Rode A, Nothnagel T, Schulz H (2016) Quantifying biochemical quality parameters in carrots (*Daucus carota* L.) – FT-Raman spectroscopy as efficient tool for metabolite profiling. Food Chem 212:495–502

Kramer M, Maksylewicz-Kaul A, Baranski R, Nothnagel T, Carle R, Kammerer DR et al (2012) Effects of cultivation year and growing location on the phenolic profile of differently coloured carrot cultivars. J App Bot Food Qual 85:235–247

Krinsky NI (1998) The antioxidant and biological properties of the carotenoids. Ann NY Acad Sci 854:443–447

Külheim C, Padovan A, Hefer C, Krause S, Köllner TG et al (2015) The *Eucalyptus* terpene synthase gene family. BMC Genomics 16:450

Kurilich AC, Clevidence BA, Britz SJ et al (2005) Plasma and urine responses are lower for acylated *vs* nonacylated anthocyanins from raw and cooked purple carrots. J Agric Food Chem 53:6537–6542

Kust AF (1970) Inheritance and differential formation of color and associated pigments in xylem and phloem of carrot, *Daucus carota*, L. PhD thesis dissertation, University of Wisconsin-Madison, USA

Laferriere L, Gabelman WH (1968) Inheritance of color, total carotenoids, alpha-carotene, and beta-carotene in carrots, *Daucus carota* L. Proc Am Soc Hortic Sci 93:408–418

Le Clerc V, Aubert C, Cottet V, Yovanopoulos C, Piquet M et al (2019) Breeding for carrot resistance to *Alternaria dauci* without compromising taste. Mol Breed 39:59

Lecomte M, Berruyer R, Hamamaa L, Boedo C, Hudhommed P et al (2012) Inhibitory effect of the carrot metabolites 6-methoxymellein and falcarindiol on development of the fungal leaf blight pathogen *Alternaria dauci*. Phys Mol Plant Path 80:58–67

Lee MW, Padilla CS, Gupta C, Galla A, Pereira A et al (2020) The *fatty acid desaturase2* family in tomato contributes to primary metabolism and stress responses. Plant Physiol 182:1083–1099

Legault J, Pichette A (2007) Potentiating effect of β-caryophyllene on anticancer activity of α-humulene, isocaryophyllene and paclitaxel. J Pharm Pharmacol 59:1643–1647

Leja M, Kamińska I, Kramer M, Maksylewicz-Kaul A, Kammerer D et al (2013) The content of phenolic compounds and radical scavenging activity varies with carrot origin and root color. Plant Foods Hum Nutr 68:163–170

Leonti M, Casu L, Raduner S, Cottiglia F, Floris C et al (2010) Falcarinol is a covalent cannabinoid CB1 receptor antagonist and induces pro-allergic effects in skin. Biochem Pharmacol 79:1815–1826

Lu S, Li L (2008) Carotenoid metabolism: biosynthesis, regulation, and beyond. J Integr Plant Biol 50:778–785

Lund ED, White JM (1990) Polyacetylenes in normal and water stressed 'Orlando Gold' carrots (*Daucus carota*). J Sci Food Agric 51:507–516

Ma J, Xu Z, Tan G, Wang F, Xiong A (2017) Distinct transcription profile of genes involved in carotenoid biosynthesis among six different color carrot (*Daucus carota* L.) cultivars. Acta Biochim Biophys Sinica 49:817–826

Maass D, Arango J, Wust F, Beyer P, Welsch R (2009) Carotenoid crystal formation in Arabidopsis and carrot roots caused by increased phytoene synthase protein levels. PLoS One 4:e6373

Maeda K, Kimura S, Demura T (2005) *DcMYB1* acts as a transcriptional activator of the carrot phenylalanine ammonia-lyase gene (*DcPAL1*) in response to elicitor treatment, UV-B irradiation and the dilution effect. Plant Mol Biol 59:739–752

Martin DM, Aubourg S, Schouwey MB, Daviet L, Schalk M et al (2010) Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. BMC Plant Biol 10:226

Mazza G, Cacace JE, Kay CD (2004) Methods of analysis for anthocyanins in plants and biological fluids. J AOAC Int 87:129–145

McCann D, Barrett A, Cooper A, Crumpler D, Dalen L et al (2007) Food additives and hyperactive behaviour in 3-year-old and 8/9-year-old children in the community: a randomised, double-blinded, placebo-controlled trial. Lancet 370:1560–1567

Meng G, Clausen SK, Rasmussen SK (2020) Transcriptome analysis reveals candidate genes related to anthocyanin biosynthesis in different carrot genotypes and tissues. Plan Theory 9:344

Milla MAR, Townsend J, Chang I, Cushman JC (2006) The Arabidopsis AtDi19 gene family encodes a novel type of *Cys2/His2* zinc-finger protein implicated in ABA-independent dehydration, high-salinity stress and light signaling pathways. Plant Mol Biol 61:13

Minto RE, Blacklock BJ (2008) Biosynthesis and function of polyacetylenes and allied natural products. Prog Lipid Res 47:233–306

Mockute D, Nivinskiene O (2004) The sabinene chemotype of essential oil of seeds of *Daucus carota* L. ssp. *carota* growing wild in Lithuania. J Essent Oil Res 16:277–281

Montilla EC, Arzaba MR, Hillebrand S, Winterhalter P (2011) Anthocyanin composition of black carrot (*Daucus carota* ssp. *sativus* var. *atrorubens* Alef.) cultivars Antonina, beta sweet, deep purple, and purple haze. J Agric Food Chem 59:3385–3390

Moreno JC, Pizarro L, Fuentes P, Handford M, Cifuentes V et al (2013) Levels of lycopene beta-cyclase 1 modulate carotenoid gene expression and accumulation in *Daucus carota*. PLoS One 8: e58144

Muchlinski A, Ibdah M, Ellison S, Yahyaa M, Nawade B et al (2020) Diversity and function of terpene synthases in the production of carrot aroma and flavor compounds. Sci Rep 10:9989

Narayan MS, Naidu KA, Ravishankar GA, Srinivas L, Venkataraman LV (1999) Antioxidant effect of anthocyanin on enzymatic and non-enzymatic lipid peroxidation. Prostaglandins Leukot Essent Fatty Acids 60:1–4

Netzel M, Netzel G, Kammerer DR, Schieber A, Carleet R (2007) Cancer cell antiproliferation activity and metabolism of black carrot anthocyanins. Innov Food Sci Emerg 8:365–372

Okuley J, Lightner J, Feldmann K, Yadav N, Lark E et al (1994) Arabidopsis FAD2 gene encodes the enzyme that is essential for polyunsaturated lipid synthesis. Plant Cell 6:147–158

Oude Griep LM, Verschuren WM, Kromhout D, Ocke MC, Geleijnse JM (2011) Colours of fruit and vegetables and 10-year incidence of CHD. Br J Nutr 106:1562–1569

Pace B, Capotorto I, Cefola M, Minasi P, Montemurro N et al (2020) Evaluation of quality, phenolic and carotenoid composition of fresh-cut purple Polignano carrots stored in modified atmosphere. J Food Compos Anal 86:103363–103370

Pateraki I, Heskes AM, Hamberger B (2015) Cytochromes P450 for terpene functionalisation and metabolic engineering. Adv Biochem Eng Biotechnol 148:107–139

Perez MB, Da Peña Hamparsomian MJ, Gonzalez RE, Denoya GI, Dominguez DLE et al (2022) Physicochemical characteristics, degradation kinetics and antioxidant capacity of anthocyanin-based aqueous extracts of purple carrots, compared to synthetic and natural colorants. Food Chem 387:132893

Perrin F, Brahem M, Dubois-Laurent C, Huet S, Jourdan M et al (2016) Differential pigment accumulation in carrot leaves and roots during two growing periods. J Agric Food Chem 64: 906–912

Perrin F, Hartmann L, Dubois-Laurent C, Welsch R, Huet S et al (2017a) Carotenoid gene expression explains the difference of carotenoid accumulation in carrot root tissues. Planta 245:737–747

Perrin F, Dubois-Laurent C, Gibon Y, Citerne S, Suel A et al (2017b) Combined *Alternaria dauci* infection and water stresses impact carotenoid content of carrot leaves and roots. Environ Exp Bot 143:125–134

Prior RL, Wu X (2006) Anthocyanins: structural characteristics that result in unique metabolic patterns and biological activities. Free Radic Res 40:1014–1028

Purup S, Larsen E, Christensen LP (2009) Differential effects of falcarinol and related aliphatic C (17)-polyacetylenes on intestinal cell proliferation. J Agric Food Chem 57:8290–8829

Radoslaw JK (2012) Hydroindene-derived chiral synthons from carotol and their cytotoxicity. Phytochemistry 23:1038–1045

Rao AV, Agarwal S (2000) Role of antioxidant lycopene in cancer and heard disease. J Am Coll Nutr 19:563–569

Reduron JP, Boulanger A, Joyeux C, Dubois-Laurent C, Huet S et al (2019) Phytochemistry of *Daucus carota* wild populations in Corsica (France). Acta Hortic 1264:157–163

Reichardt S, Budahn H, Lamprecht D, Riewe D, Ulrich D et al (2020) The carrot monoterpene synthase gene cluster on chromosome 4 harbours genes encoding flavour-associated sabinene synthases. Hortic Res 7:190

Rodriguez-Concepcion M, Stange C (2013) Biosynthesis of carotenoids in carrot: an underground story comes to light. Arch Biochem Biophys 539:110–116

Santos C, Simon PW (2002) QTL analyses reveal clustered loci for accumulation of major provitamin A carotenes and lycopene in carrot roots. Mol Gen Genet 268:122–129

Santos C, Simon PW (2006) Heritabilities and minimum gene number estimates of carrot carotenoids. Euphytica 151:79–86

Santos C, Senalik D, Simon PW (2005) Path analysis suggests phytoene accumulation is the key step limiting the carotenoid pathway in white carrot roots. Genet Mol Biol 28:287–293

Santos P, Busta L, Yim WC, Cahoon EB, Kosma DK (2022) Structural diversity, biosynthesis, and function of plant falcarin-type polyacetylenic lipids. J Exp Bot. https://doi.org/10.1093/jxb/erac006

Schmid C, Sharma S, Stark TD, Günzkofer D, Hofmann TF et al (2021) Influence of the abiotic stress conditions, waterlogging and drought, on the bitter sensometabolome as well as agronomical traits of six genotypes of *Daucus carota*. Foods 10:1607

Schmiech L (2010) Strukturaufklärung und quantitative Studien zu Bitterstoffen und Polyacetylenen in Karotten (*Daucus carota* L.). PhD dissertation thesis, Technische Universität München, München, Germany

Schmiech L, Alayrac C, Witulski B, Hofmann T (2009) Structure determination of bisacetylenic oxylipins in carrots (*Daucus carota* L.) and enantioselective synthesis of falcarindiol. J Agric Food Chem 57:11030–11040

Schulz-Witte J (2011) Diversität wertgebender Inhaltsstoffe bei *Daucus carota* L. PhD dissertation thesis, Technical University Braunschweig, Braunschweig, Germany

Semba RD (1998) The role of vitamin A and related retinoids in immune function. Nutr Rev 56: s38–s48

Sevimli-Gur C, Cetin B, Akay S, Gulce-Iz S, Yesil-Celiktas O (2013) Extracts from black carrot tissue culture as potent anticancer agents. Plant Foods Hum Nutr 68:293–298

Shaltiel-Harpaz L, Yahyaa M, Nawade B, Dudareva N, Ibdah M (2021) Identification of a wild carrot as carrot psylla (*Bactericera trigonica*) attractant and host plant chemistry. Plant Sci 311: 111011

Sharma KD, Karki S, Thakur NS, Attri S (2012) Chemical composition, functional properties and processing of carrot – a review. J Food Sci Technol 49:22–32

Shebaby WN, El-Sibai M, Bodman-Smith K, Karam MC, Mroueh M et al (2013) The antioxidant and anticancer effects of wild carrot oil extract. Phytother Res 27:737–744

Shirley BW (1996) Flavonoid biosynthesis: "new" functions for an "old" pathway. Trends Plant Sci 1:377–382

Sieniawska E, Swiatek L, Rajtar B, Kozioł E, Polz-Dacewicz M et al (2016) Carrot seed essential oil – source of carotol and cytotoxicity study. Ind Crop Prod 92:109–115

Simon PW (1996) Inheritance and expression of purple and yellow storage root color in carrot. J Hered 87:63–66
Simon PW (2000) Domestication, historical development, and modern breeding of carrot. Plant Breed Rev 19:157–190
Simon PW, Wolff XY (1987) Carotenes in typical and dark orange carrots. J Agric Food Chem 35: 1017–1022
Simon PW, Wolff XY, Peterson CE, Kammerlohr DS, Rubatzky VE et al (1989) High carotene mass carrot population. HortScience 24:174
Simon PW, Geoffriau E, Ellison S, Iorizzo M (2019) Carrot carotenoids genetics and genomics. In: Simon PW, Iorizzo M, Grzebelus D, Baranski R (eds) The carrot genome. Springer, Cham, pp 247–260
Simpson K, Cerda A, Stange C (2016) Carotenoid biosynthesis in *Daucus carota*. In: Stange C (ed) Carotenoids in nature, Subcellular biochemistry, vol 79. Springer, Cham, pp 199–217
Singh DP, Beloy J, McInerney JK, Day L (2012) Impact of boron, calcium and genetic factors on vitamin C, carotenoids, phenolic acids, anthocyanins and antioxidant capacity of carrots (*Daucus carota*). Food Chem 132:1161–1170
Singleton VL, Orthofer R, Lamuela-Raventos RM (1999) Analysis of total phenols and other oxidation substrates and antioxidants by means of Folin–Ciocalteu reagent. Methods Enzymol 299:152–178
Søltoft M, Nielsen J, Holst Laursen K, Husted S, Halekoh U et al (2010) Effects of organic and conventional growth systems on the content of flavonoids in onions and phenolic acids in carrots and potatoes. J Agric Food Chem 58:10323–10329
Soto-Vaca A, Gutierrez A, Losso JN, Xu Z, Finley JW (2012) Evolution of phenolic compounds from color and flavor problems to health benefits. J Agric Food Chem 60:6658–6677
Stange Klein C, Rodriguez-Concepcion M (2015) Carotenoids in carrots. In: Chen C (ed) Pigments in fruits and vegetables. Springer, New York, pp 217–228
Stange C, Fuentes P, Handford M, Pizarro L (2008) *Daucus carota* as a novel model to evaluate the effect of light on carotenogenic gene expression. Biol Res 41:289–301
Stevenson D, Hurst R (2007) Polyphenolic phytochemicals – just antioxidants or much more? Cell Mol Life Sci 64:2900–2916
Sun T, Simon PW, Tanumihardjo SA (2009) Antioxidant phytochemicals and antioxidant capacity of biofortified carrots (*Daucus carota* L.) of various colors. J Agric Food Chem 57:4142–4147
Sun T, Yuan H, Cao H, Yazdani M, Tadmor Y et al (2018) Carotenoid metabolism in plants: the role of plastids. Mol Plant 11:58–74
Surjadinata BB, Cisneros-Zevallos L (2017) Biosynthesis of phenolic antioxidants in carrot tissue increases with wounding intensity. Food Chem 134:615–624
Surjadinata BB, Jacobo-Velázquez DA, Cisneros-Zevallos L (2017) UVA, UVB and UVC light enhances the biosynthesis of phenolic antioxidants in fresh-cut carrot through a synergistic effect with wounding. Molecules 22:668
Sylvestre-Gonon E, Law SR, Schwartz M, Robe K, Keech O et al (2019) Functional, structural and biochemical features of plant serinyl-glutathione transferases. Front Plant Sci 10:608
Tan KW, Killeen DP, Li Y, Paxton JW, Birch NP et al (2014) Dietary polyacetylenes of the falcarinol type are inhibitors of breast cancer resistance protein (BRCP/ABCG2). Eur J Pharmacol 723:346–352
Tanumihardjo SA (ed) (2012) Carotenoids and human health. Springer, New York, 331 p
Tavares AC, Gonçalves MJ, Cavaleiro C, Cruz MT, Lopes MC et al (2008) Essential oil of *Daucus carota* subsp. *halophilus*: composition, antifungal activity and cytotoxicity. J Ethnopharmacol 119:129–134
Tholl D (2006) Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. Curr Opin Plant Biol 9:297–304
Ulrich D, Nothnagel T, Schulz H (2015) Influence of cultivar and harvest year on the volatile profiles of leaves and roots of carrots (*Daucus carota* spp. *sativus* Hoffm.). J Agric Food Chem 63:3348–3356

Umiel N, Gabelman WH (1972) Inheritance of root color and carotenoid synthesis in carrot, *Daucus carota*, L.: orange *vs.* red. J Am Soc Hortic Sci 97:453–460

Vasudevan M, Gunnam KK, Parle M (2006) Antinociceptive and anti-inflammatory properties of *Daucus carota* seeds extracts. J Health Sci 52:598–606

Vivek BS, Simon PW (1999) Linkage relationships among molecular markers and storage root traits of carrot (*Daucus carota* L. ssp. *sativus*). Theor Appl Genet 99:58–64

Wako T, Kimura S, Chikagawa OY (2010) Characterization of MYB proteins as transcriptional regulatory factors for carrot phenylalanine ammonia-lyase gene (*DcPAL3*). Plant Biotechnol 27: 131–139

Wang H, Ou CG, Zhuang FY, Ma ZG (2014) The dual role of phytoene synthase genes in carotenogenesis in carrot roots and leaves. Mol Breed 34:2065–2079

Warner SR (2019) Carrots and cancer: the bioavailability of polyacetylene from carrots and their association with biomarkers of cancer risk. PhD thesis dissertation, Human Nutrition Research Centre, Faculty of Medical Sciences, Newcastle University, UK

Xu ZS, Huang Y, Wang F (2014) Transcript profiling of structural genes involved in cyanidin-based anthocyanin biosynthesis between purple and non-purple carrot (*Daucus carota* L.) cultivars reveals distinct patterns. BMC Plant Biol 14:262

Xu ZS, Ma J, Wang F, Ma HY, Wang QX et al (2016) Identification and characterization of *DcUCGalT1*, a galactosyltransferase responsible for anthocyanin galactosylation in purple carrot (*Daucus carota* L.) taproots. Sci Rep 6:27356

Xu Z-S, Feng K, Que F, Wang QX, Xiong A-S (2017) A MYB transcription factor, *DcMYB6*, is involved in regulating anthocyanin biosynthesis in purple carrot taproots. Sci Rep 7:45324

Xu Z-S, Yang Q-Q, Feng K, Xiong A-S (2019) Changing carrot color: insertions in *DcMYB7* alter the regulation of anthocyanin biosynthesis and modification. Plant Physiol 181:195–207

Xu Z-S, Yang Q-Q, Feng K, Yu X, Xiong AS (2020) *DcMYB113*, a root-specific R2R3-MYB, conditions anthocyanin biosynthesis and modification in carrot. Plant Biotechnol J7:1585–1597

Yahyaa M, Tholl D, Cormier G, Jensen R, Simon PW, Ibdah M (2015) Identification and characterization of terpene synthases potentially involved in the formation of volatile terpenes in carrot (*Daucus carota* L.) roots. J Agric Food Chem 63:4870–4878

Yahyaa M, Ibdah M, Marzouk S, Ibdah M (2016) Profiling of the terpene metabolome in carrot fruits of wild (Daucus carota L. ssp. carota) accessions and characterization of a geraniol synthase. J Agric Food Chem 66:2378–2386

Yildiz M, Willis DK, Cavagnaro PF, Iorizzo M, Abak K et al (2013) Expression and mapping of anthocyanin biosynthesis genes in carrot. Theor Appl Genet 126:1689–1702

Yoshitomi K, Taniguchi S, Tanaka K, Uji Y, Akimitsu K et al (2016) Rice terpene synthase 24 (OsTPS24) encodes a jasmonate responsive monoterpene synthase that produces an antibacterial γ-terpinene against rice pathogen. J Plant Physiol 191:120–126

Yuan H, Zhang J, Nageswaran D, Li L (2015) Carotenoid metabolism and regulation in horticultural crops. Hortic Res 2:15036

Zaini RG, Brandt K, Claench MR, Le Maitre CL (2012) Effects of bioactive compounds from carrots (*Daucus carota* L.), polyacetylenes, beta-carotene and lutein on human lymphoid leukaemia cells. Anti Cancer Agents Med Chem 12:640–652

Zhang D, Hamauzu Y (2004) Phenolic compounds and their antioxidant properties in different tissues of carrots (*Daucus carota* L.). J Food Agric Environ 2:95–100, ISSN:14590263

Zhang Y, Butelli E, Martin C (2014) Engineering anthocyanin biosynthesis in plants. Curr Opin Plant Biol 19:81–90

Zhang H, Liu R, Tsao R (2016) Anthocyanin-rich phenolic extracts of purple root vegetables inhibit pro-inflammatory cytokines induced by H_2O_2 and enhance antioxidant enzyme activities in Caco-2 cells. J Funct Foods 22:363–375

Zidorn C, Jöhrer K, Ganzera M, Schubert B, Sigmund EM et al (2005) Polyacetylenes from the Apiaceae vegetables carrot, celery, fennel, parsley, and parsnip and their cytotoxic activities. J Agric Food Chem 53:2518–2525

Metabolomics and Cytoplasmic Genomics of *Allium*

Mostafa Abdelrahman, Rawan Rabie, Magdi El-sayed, and Masayoshi Shigyo

Contents

Abstract

Allium is the biggest genus of petaloid monocotyledons, with more than 750 species widely distributed in a range of climatic conditions worldwide, especially in the Northern Hemisphere. *Allium* comprises commercially significant food crops, including onions, garlic, leeks, and chives, as well as species with medicinal

M. Abdelrahman (✉)
Aswan University Faculty of Science, Aswan, Egypt
e-mail: moustafa.abdelwahd@Gu.edu.eg

R. Rabie · M. El-sayed
Faculty of Science, Galala University, Suze, Egypt

M. Shigyo
Laboratory of Vegetable Crop Science, College of Agriculture, Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Yamaguchi, Japan

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_52

qualities. Bulb onions rank second only after tomatoes in terms of global production, which indicates the importance of *Allium* crops and the need for developing new *Allium* crop varieties with beneficial agronomical traits. Recently, there has been considerable interest in investigating the genetic resources of *Allim* crops and their wild relatives for improving *Allium* breeding and possible future genetic manipulation. This chapter provides a comprehensive review of major *Allium* crops and their wild relatives from scientific and horticultural perspectives. This chapter broadly covers the unique resources for *Allium* genetics and breeding, including the recent development of cytoplasmic male sterility, inbred lines, and wild species. We also discuss and summarize the recent developments in *Allium* genome sequencing, including novel tools for large genome sequencing, the chloroplast genome, mitochondrial genomes, and the nuclear genome. Furthermore, we provide a brief overview of the linkage, cytogenetic, and physical mapping in various Allium crops. Finally, we provide a special focus on *Allium* metabolome and transcriptome analysis as important approaches for understanding *Allium* stress responses. Our book chapter provides recent developments in *Allium* genomics and metabolome dynamics, which open the possibility of developing novel *Allium* crop cultivars with enhanced nutritional value and stress tolerance under current climatic conditions.

Keywords

Allium · Cytogeneitcs · *Allium* genome · Genetic resource · *Allium* metabolomics · Stress tolerance

1 Introduction

Onions (*A. cepa* L. Common onion group), shallots (*A. cepa* L. Aggregatum group), garlic (*A. sativum* L.), Japanese bunching onion (*A. fistulosum* L.), and leeks (*A. ampeloprasum* L. Leek group) are all in the same genus *Allium* that have been cultivated globally as important spices, condiments, and vegetables (Abdelrahman et al. 2016). The genus *Allium* comprises ~1000 species (Govaerts et al. 2005–2020), of which onions, shallots, and leeks are widely cultivated, and displayed remarkable diversity in their morphological and physiological properties, particularly in life form (Abdelrahman et al. 2017a, b, c). *Allium* is propagated commonly through the Northern Hemisphere from the boreal region to the dry subtropics. A region with high density of species diversity started from the Mediterranean Basin to Central Asia, in addition to a second smaller centre of species diversity which is located in North America (Fritsch et al. 2010; Abdelrahman et al. 2017a, b, c) (Fig. 1).

Additionally, *Allium* species show significant variation in a number of cytogenetic traits, including ploidy level (ranging from 2× to 16×), genome size, and chromosomal numbers ($x = 7$, 8, and 9). Specifically, the largest genomes have been found in Alliums, ranging from 18.1 picograms (pg) in *A. triquetrum* L. to 31.5 pg in *A. ursinum* L. (Bennett et al. 2000; Peška et al. 2019). For instance, the haploid

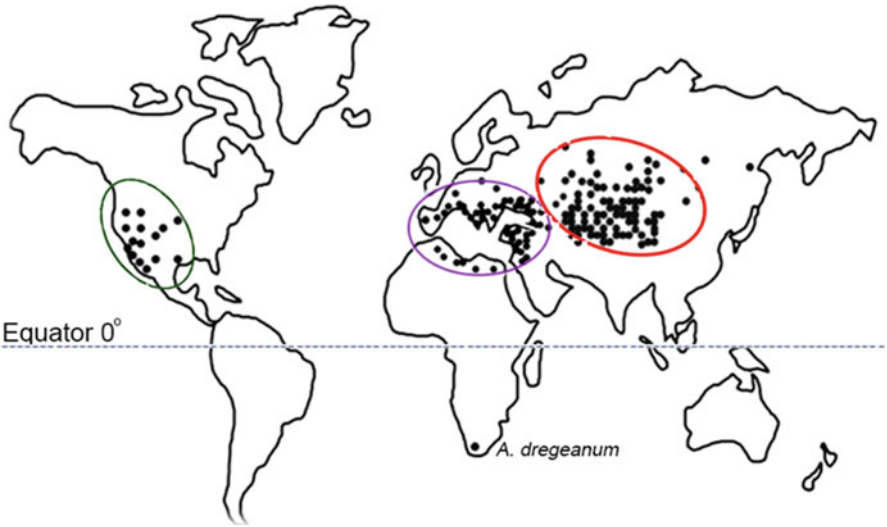


Fig. 1 *Allium* distribution in the northern and southern hemisphere. The red and purple circles indict the center for high *Allium* diversity from Central Asia to the Mediterranean Basin

genome size of the bulb onion was calculated to be about 16,400 Mega bases (Mb)/1C, making it the largest among all cultivated diploid crops. This size is 120 times larger than *Arabidopsis* (ca. 125 Mb) and 5 times larger than the human genome, which is only about 3000 Mb (Ricroch et al. 2005). At least 95% of the onion genome is made up of repetitive sequences (Vitte et al. 2013), which are typically associated with large genome sizes (Kelly and Leitch 2011). As a result, mapping and genomic-assisted breeding in Alliums are behind that of other crops like tomato (*Solanum lycopersicum*), wheat (*Triticum aestivum*), and rice (*Oryza sativa*), making the entire genome sequencing of bulb onions a considerable problem. With the development of next-generation sequencing (NGS) technology, DNA sequencing of many plant species, including *Allium*, grew faster and more efficient with higher throughputs and greater genome coverage (Abdelrahman et al. 2017a, b, c; Abdelrahman 2020; Valliyodan et al. 2017). With the use of these NGS technologies, the first waves of crop genome sequencing, gene expression atlases, and a better knowledge of the signalling networks underlying plant responses to biotic and abiotic stresses were all made possible (Rothberg et al. 2011; Abdelrahman et al. 2019). In this chapter, we report the first trail of the genome assembly of doubled haploid (DH) *A. cepa*. Additionally, the presentation of the *Allium* genome organization, which includes the nuclear, mitochondrial, and chloroplast genomes are given. This chapter also addressed how the genetic resources of the *Allium* species, such as DHs, alien chromosomal addition lines, and cytoplasmic substitution lines, have been employed in various genetic research to comprehend the discrepancy between *Allium*'s phenotype and genotype traits. This chapter's contents will be a useful tool for studying onion breeding genetic resources, which are valuable for onion breeding and research.

2 Unique Resources for Genetics and Breeding in *Allium*

Since Peffley et al. (1985) published their pioneering research, monosomic addition lines (MALs) have been useful genetic resources for studying important agronomic traits in cultivated Alliums and assigning linkage groups to physical chromosomes. For example, in a backcross population (BC_2) of amphidiploid hybrids between *A. fistulosum* (FF) and shallot (*A. cepa* Aggregatum group, AA), a full set of *A. fistulosum* and *A. cepa* MALs ($2n = 2x + 1 = 17$, genomes + chromosomes FF +1A-FF+8A) was identified (Shigyo et al. 1996). To confirm the integrity of the chromosomal constitutions, the MALs were subjected to genomic in situ hybridization (GISH) as described by Shigyo et al. (1996). Following that, 48 shallot chromosome-specific molecular markers were developed in conjunction with qualitative data for morphological identifications (Shigyo et al. 1997a, b). All of these morphological characteristics appear to be related to the genetic effects of the shallot additional chromosome(s) on the A. fistulosum integral diploid background. The reddish-brown sheaths of FF+5A indicated that this line had a high flavonoid and anthocyanin content (Shigyo et al. 1997b). Evidently, integrated metabolome and transcriptome assessment of the entire set of MALs provides additional evidence that shallot chromosome 5A harbored many candidate genes associated in flavonoid biosynthesis (Abdelrahman et al. 2019). When compared to other MALs, FF+5A exhibited high expression of many downstream and upstream flavonoid and anthocyanin-related genes, which was coherent with the accumulation of several flavonoids and anthocyanin-related compounds. Similarly, saponin transcriptome and phytochemical screening in MALs revealed that FF+2A accumulated Alliospiroside, a saponin with powerful antifungal activity against Fusarium pathogens and an useful attribute for disease resistance (Abdelrahman et al. 2017a, b, c). Furthermore, Yaguchi et al. (2013) also investigated antioxidant capacity and polyphenol concentration in MALs, discovering that the FF+2A and FF+6A lines had the greatest polyphenol concentration and antioxidant capacity. As a result, the additional chromosomes 2A and 6A contain anonymous genes associated with polyphenol production increased expression. Shigyo and co-workers carried out such approaches in six interspecific-combinations, *A. cepa – A. fistulosum* (Hang et al. 2004; Yaguchi et al. 2008), *A. cepa – A. roylei* Stearn (Vu et al. 2012), *A. fistulosum– A. roylei* (Ariyanti et al. 2015), *A. fistulosum – A. galanthum* Kar. et Kir. (Tashiro et al. 2000), *A. fistulosum – A. oschaninii* O. Fedtsch. (Tashiro et al. 2000), and *A.fistulosum – A. vavilovii* M. Pop. et Vved. (Tashiro et al. 2000) together with further approaches for disomic additions (Shigyo et al. 2003; Yaguchi et al. 2008) and multiple alien chromosome additions (Masuzaki et al. 2007), some of which are relevant to elucidating a specific epistatic regulation related to secondary metabolite biosynthesis (Masuzaki et al. 2006).

2.1 Cytoplasmic Male Sterility (CMS)

CMS is a maternally inherited trait that prevents the formation of viable pollens while leaving female gametes unaffected, allowing them to serve as seed parents in hybrid

seed production (Yamashita et al. 2007). In contrast, a single nuclear locus renowned as *Ms* is known to revive CMS male fertility in onions. In a recent study by Yu and Kim (2020), inheritance patterns in segregating populations implied that the *Ms* and *Ms2* loci were both engaged in fertility restoration in onions. The *Ms2* locus was found at the end of chromosome 2 at a distance of 70cM from the *Ms* locus (Yu and Kim 2020). Although it is unknown whether the causal genes for *Ms* and *Ms2* loci are paralogs, transcription levels of *orf725*, a CMS-associated gene in onions, were significantly down-regulated in male-fertile individuals of segregating populations, suggesting that the causal gene for *Ms2* may be able to restore male fertility by suppressing transcription of *orf725* or degrading transcripts (Yu and Kim 2020).

Allium galanthum (G) cytoplasm induces CMS in bulb onion (Havey 2002), Japanese bunching onion (Yamashita et al. 1999a, b, 2002, 2005; Yamashita and Tashiro 2004), and shallot (Yamashita et al. 1999). In shallot, the G cytoplasm was introduced through continuous backcrossing to shallot the recurrent parent, to obtain interspecific F_1 hybrids (Yamashita et al. 1999). *A. roylei* (R) originated CMS lines, were identified in the BC_2 progenies of a single amphidiploid (possessing Rcytoplasm) between *A. roylei* (female parent) and *A. cepa*. Breeding onion F_1 hybrid varieties takes 10–12 years to establish the A (malesterile line), B (male-sterile maintainer), and C (restorer) lines (Shigyo and Kik 2008), while alloplasmic lines possessing either GorR cytoplasm saves breeders' time and nullifies the need for B line. Moreover, since the CMS system is conditioned by the incompatibility between the wild species' cytoplasm and the *A. cepa* nucleus, any onion population acts as a male sterility maintainer. Unfortunately, no G or R CMS onion lines have been used in actual onion F_1 seed production. It should be noted, however, that compared to the S-type cytoplasmic male sterility, there are concerns about nectar production, a preferred factor by flower-visiting insects, and/or interaction of pollen production with the environment. Further research on the ecology and physiology of cytoplasm in wild species is needed.

2.2 Inbred Lines

The traditional breeding of onion inbred lines is long and difficult because of its biennial life span, inbreeding depression, and high heterozygosity. Thus, DH lines provide several advantages, including complete homozygosity, reduced DNA methylation, and elimination of deleterious alleles compared with inbred lines (Bohanec 2002) and consequently exhibit vigorous vegetative growth, morphological uniformity (Hyde et al. 2012; Khan et al. 2020). Single- and two-step protocols have been used to generate gynogenic haploids in onion. The former involves culturing whole flower buds, ovaries, or ovules to the embryo, or plantlet stage (Bohanec and Jakše 1999; Bohanec 2002). Whereas the latter includes preculture of flower buds on basal media with or without plant growth regulators. Then, isolation of ovary or ovule and subsequent subculturing on regeneration media with growth regulators (Michalik et al. 2000; Martinez et al. 2000). DH lines may be induced in vitro by culturing haploid cells from the male (androgenesis) or female (gynogenesis) gametophytes. However, the efficiency of androgenic generation DHs is limited in some *Allium*

cepa genotypes due to defective pollen or anthers, or/and the production of albinos during culture. Gynogenesis has been the most commonly used method for producing haploids in *A. schoenoprasum* L. (Keller 1990), leek (Kaska et al. 2013; Alan et al. 2016), *A. altaicum* (Keller 1990), *A. giganteum* (Susek et al. 2002), shallot (Sulistyaningsih et al. 2006), *A. fistulosum* (Ibrahim et al. 2016), *A. tuncelianum* (Yarali and Yanmaz 2016), and others.

DH lines offer very promising tools to understand the gap between phenotype and genotype, and DH chromosomal doubling can shorten the time, offer homozygous pure lines, and provide valuable materials for genomic analysis. For example, comparative metabolome analysis of shallot and bulb onion DHs (DHA, and DHC, respectively) and F_1 hybrid revealed genotype-specific metabolites for each. DHA accumulated several stress-related metabolites, which is consistent with the biotic and abiotic stress shallot tolerance compared with bulb onion (Abdelrahman et al. 2015). Similarly, Khosa et al. (2016) developed transcriptome data from reproductive and vegetative organs of onion DH line 'CUDH 2107' to identify tissue-specific expressed genes involved in pollen fertility.

The Cornell onion breeding program compared 20 DH lines with commercial hybrids and open-pollinated cultivars (OP) developed from the same source germplasm over 2 years, under field conditions (Hyde et al. 2012). Results indicated a similar vegetative vigor of both DH and OPs, with minimal inbreeding depression. A comparison between two sets of hybrids produced using male DH lines and two different commercial females showed increased vegetative growth vigor of DH lines and their derived hybrids probably due to elimination of deleterious lethal genes during gynogenesis, an additional benefit for onion breeding strategies (Hyde et al. 2012).

DH lines lack allelic exchanges within a locus, thus genetic effects are controlled by single alleles or interactions between *loci*, thus facilitating detection of quantitative trait loci (QTL) (Duangjit et al. 2014). Hence, onion DH lines from the four Spanish cultivars, including highly pungent landrace 'BGHZ1354', sweet cultivar 'Fuentes de Ebro', and 'Recas' and 'Rita,' two commercial Valenciana-type varieties (Fayos et al. 2015) served for the development of linkage maps. More comprehensive research and improved DH in vitro gynogenesis production methods are needed and would be useful to produce DH in onion and other *Allium* species.

2.3 Wild Species

Wild relatives have great potential to expand the supply of usable genetic variation and useful traits, thus are important for introgression breeding (Hao et al. 2020). *Allium roylei* (RR, $2n = 16$) possess several desired traits, including resistance to Botrytis leaf blight and downy mildew (Abdelrahman et al. 2014). Similarly, *Alliumfistulosum* (AF) exhibits several important agronomical traits for onion improvement (Matsuse et al. 2022), thus great attention has been given to the introgression of *A. fistulosum*genes to *A. cepa*. However, the interspecific hybrid is largely sterile, and heteromorphic bivalents are present (Emsweller and Jones 1935; Albini and Jones 1990). In addition, the backcross (BC_1) of the interspecific hybrid

with *A. cepa* usually yields a small progeny, due to nucleo-cytoplasmic incompatibility (van der Valk et al. 1991). Using *A. roylei* as a bridging species helps to circumvent these problems. Hence, *A. cepa* (CC) × *A. roylei* (CR), *A. cepa* × *A. fistulosum* (CF), and *A. fistulosum* × *A. roylei* (FR) (Khrustaleva and Kik 1998) were bridge-crossed with (CC × FR) where *A. Cepa* served as the female parent and *A. roylei* as a male parent (Khrustaleva and Kik 1998). Thereafter, genome organization of the interspecific hybrids and trihybrid population were analyzed using multicolor GISH (Khrustaleva and Kik 1998).

In a follow-up study, trihybrid genotypes from the CC × RF population where only one homolog of a chromosome pair underwent interspecific recombination, were used for GISH analysis of chromosomes 5 and 8 recombinants. Visualization of recombination points and the physical positions of recombination were integrated into amplified fragment length polymorphism (AFLP) linkage maps of both chromosomes, thus concluding that in *Allium,* recombinations predominantly occur in the proximal half of chromosome arms. Of the *Pst*I/*Mse*I markers, 57.9% are located near the centromeric region, thus suggesting genes' presence in this region (Khrustaleva et al. 2005).

Investigations of trihybrid 'CC × RF' population for root traits inheritance revealed quantitative trait loci (QTLs) for the rooting system, implying that breeding for an enhanced rooting system in onion is feasible (De Melo 2003). Analyzes of trihybrid 'CC × CF' population for the genetic background for response to arbuscular mycorrhizal fungus *Glomus intraradices* revealed that the amalgamation of three genomes increases the genetic variation for plant development and the mycorrhizal reaction (Galván Vivero et al. 2011). On linkage group 9, one QTL for the number of stem-borne roots of mycorrhizal plants was discovered, which was connected with *A. fistulosum* alleles. This QTL co-segregated with QTLs for mycorrhizal and nonmycorrhizal plant average performance and total dry weight. Another QTL for Fusarium basal rot (FBR) resistance from *A. roylei* was identified on a distal region of chromosome 2, and one QTL from *A. fistulosum*was identified on the long arm of chromosome 8 (Galván Vivero 2009). These QTLs showed an additive effect and together accounted for 31 and 40% of the total variation for FBR incidence and severity at harvest; and 31 and 29% after storage, respectively.

3 Genomes

Plant cells have three genomes, mitochondrial, plastid, and nuclear. The number of genome copies per organelle depends on the cell type, the age of tissue, and species. Hence, each *Arabidopsis thaliana* (L.) Heynh cell has two copies of the nuclear genome (gDNA) (Lutz et al. 2011). The number of mitochondria and plastid DNAs (mtDNA, and ptDNA, respectively), however, vary with cell types. *Arabidopsis*root cells have ~400 mtDNA (Kato et al. 2008), whereas maize anther cells have a 20 to 40-fold higher number of mtDNA (Warmke and Lee 1978). In *Arabidopsis* and sugar beet, the copy number of ptDNAs remains at ~1700 per a single gDNA molecule, and the ratio of ptDNA to gDNA remains constant even as the ploidy level of the cell changes (Lutz et al. 2011).

3.1 Chloroplast Genome

The chloroplast (cp), which evolved from photosynthetic bacteria, is the primary organelle responsible for carbon fixation and photosynthesis in green plants (Cho et al. 2015). Most angiosperms have maternally inherited cps genomes and genetic information (Song et al. 2017). The evolutionary conserved genome architecture and gene content assisted in the creation of genetic markers for DNA barcoding, molecular identification, phylogenetic classification, and genetic resource and breeding screening. (Hu et al. 2016). The *Allium* cp whole-genome DNA sequencing initiation (von Kohn et al. 2013) resulted in cpgenomes sequencing of *A. obliquum* L., *A. prattii* C. H. Wright, *A. victorialis* L. Sp. Pl. 1: 295. 1753, *A. cepa*, and *A. sativum* (http://www.ncbi.nlm.nih.gov/genome/organelle/). Recently, the cp genomes of *A. tuberosum* Rottl. ex Spreng., *A. fistulosum*, *A. sativum,* and *A. cepa* have been sequenced using NGS technologies (Huo et al. 2019), and comparisons with cp genomes of *A. cepa* CMS-T and CMS-S, *A. victorialis, A. obliquum*, and *A. prattii* revealed that the length of cp DNAs ranged from 152 (*A. obliquum*) to 154 kb (*A. prattii*) and that the genomic structure and gene organization are well conserved (Huo et al. 2019).

The *Allium* cp genome contained large and small single-copy regions (LSC and SSC, respectively) separated by large inverted repeats (IRs) (Fig. 2). The nine cp genomes had typically quadripartite structures, with two IRs (ranged from 26,370 to 26,564 bp) splitted by the LSC (81,588 to 83,392 bp) and SSC (17,853 to 18,066 bp) regions (Huo et al. 2019) (Fig. 2).

Furthermore, all Allium cp genomes had a similar gene arrangement and content, with 140–141 genes in the IR regions, including 8 rRNA genes, 5–10 pseudogenes, 37–38 tRNA genes, and 88–89 protein-coding genes (Huo et al. 2019) (Fig. 2). The GC copy number of nine complete cp genomes was very similar in each region, including the entire cp genome, LSC, IR, SSC, coding sequences (CDSs), rRNA, tRNA, and pseudogene. The highest was observed in rRNA (>55%) and the lowest in SSC (>29%). The numbers and distributions of 3 repeat types in the 9 *Allium* cp genomes, including 394 repeats with 131 tandem repeats, 154 dispersed repeats, and 109 palindromic were similar and conserved. The number of simple sequence repeats (SSRs) varies between the nine *Allium* genomes, ranging from 73 in *A. tuberosum* to 96 in *A. cepa*. CMS-N and CMS-T (Huo et al. 2019). In the phylogentic analysis of the genus *Allium*, nine accessions were separated into two sister clusters. The first cluster comprised *A. prattii* and *A. victorialis*; the second cluster comprised seven entrees, in which *A. cepa* CMS-T and CMS-N were grouped in a sister branch and then clustered with *A. fistulosum*, *A. obliquum*, *A. cepa* CMS-S, *A. sativum*, and *A. tuberosum* (Huo et al. 2019).

3.2 Mitochondrial Genome

Scheffler (2008) stipulates that the plant mitochondrial genome harbors important information, such as mode of gene expression, organizational diversity, evolution, and nuclear-cytoplasmic interactions. Collectively these make the mitochondrial

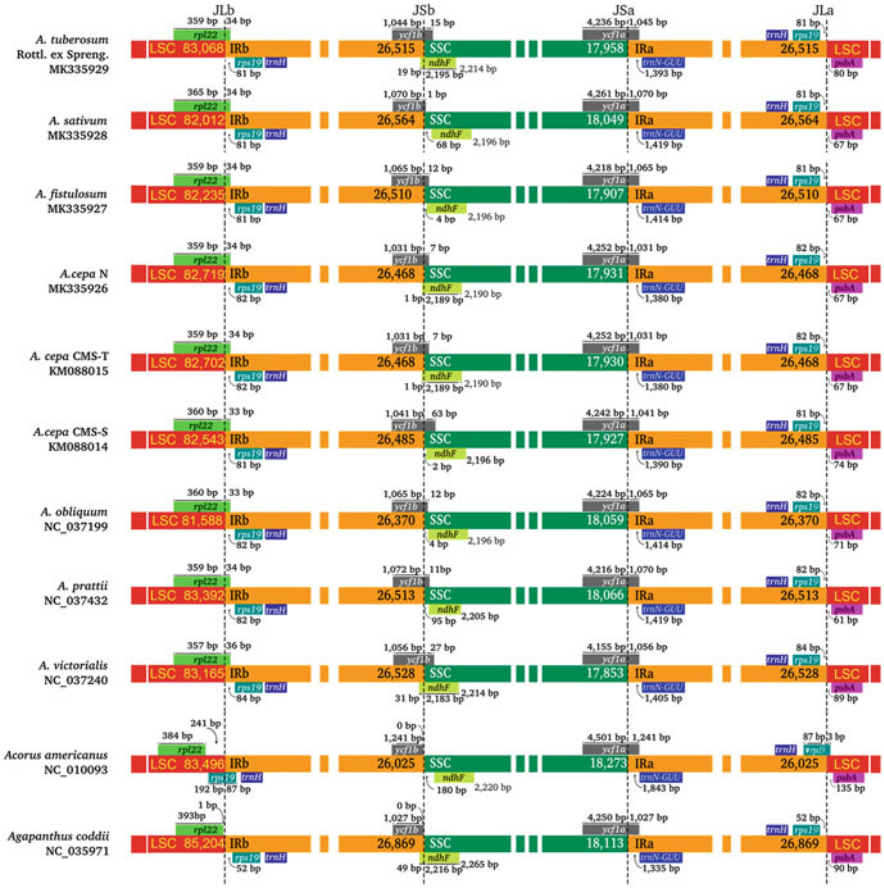


Fig. 2 Comparison of the large single copy (LSC), inverted repeat (IR), and small single-copy (SSC) boundary regions among the nine chloroplast genomes of different *Allium* species. The numbers indicate the distances from the end of the gene to the boundary sites. (This figure is adopted from Huo et al. 2019 after modification)

genome one of the most interesting genomes to molecular biologists and plant genetics. Plant mitochondrial genomes typically have many long and short repeated sequences and intra- and intermolecular recombination may create various DNA molecules in this organelle, and recombination may create a novel gene that causes CMS (Tsujimura et al. 2019). The complete onion mitochondrial genome sequence was reported for a CMS-S inbred line (Kim et al. 2016), of which ~10% consisted of repetitive sequences with short repeats of <100 base pairs. Additionally, the gene encoding cytochrome c biogenesis protein ($ccmF_N$) was split into two genes (Kim et al. 2016), as recorded in 30 other *Allium* species [REF]. In a follow-up study, comparative analysis of mitochondrial genome sequences of two recently diverged cytoplasms CMS-S and CMS-T-like resulted in the identification of *orf725*, a

chimeric gene consisting of *cox1* with other sequences, as the casual gene responsible for CMS in onions (Kim et al. 2018). In addition, CMS-T-like cytoplasm has recently diverged from the normal cytoplasm by the introduction of *orf725* (Kim et al. 2018). More recently, NGS sequencing of the mitochondrial genome of the bulb onion variety 'Momiji-3' possessing CMS-S-type cytoplasm revealed that the three circles as a result of recombination. In addition, mitochondrial genome structure was visualized using pulsed-field gel electrophoresis (Tsujimura et al. 2019). Further, mapping the transcript data to mitochondrial genome indicated that a new functional gene was absent from the 'Momiji-3' mitochondrial genome f and confirmed *orf725* candidacy gene for CMS. Recently, Hui et al. (2020) published the complete sequence of the onion's yellow stripe mutant mitochondrial genome. Compared with normal onions, 556 single nucleotide polymorphisms (SNPs), 279 small indel numbers were detected, out of which 14 SNPs and two small indels in the exon region affected the translation protein. It should be noted, however, that studies on purified *Allium* mitochondrial DNA have not yet been performed thus the number of differences between onion genotypes remains unclear.

3.3 Nuclear Genome

A study of *Allium* gDNA in 42 indicated that 4C nuclear DNA contents ranged between 41.19 and 142.78 pg, irrespective of the basic chromosome number, ploidy, or taxonomic group (Labani and Elkington 1987). Similarly, there was no correlation between the proportion of the C-banded region in the karyotype and the increase in DNA nuclear content. For example, *A. ursinum* L. has the largest DNA content of 71.39pg/4C DNA/genome but showed the lowest proportions of C-banded chromosomal region in the genus, while *Allium* species with moderate DNA amounts (21.11–42.45 pg)/4C DNA/genome showed the highest proportions of C-banded chromosomal region (Vosa 1976; Al-Sheikh Hussain 1977).

In general, DNA amounts in *Allium* species showed high divergence. For instance, the DNA values (2C) in triploid *A. carinatum* L. vary between 32 $\pm$ 67 and 50 $\pm$ 28 pg (Nagl and Fusenig 1979; Labani and Elkington 1987) and in leek between 24 $\pm$ 1 and 60 $\pm$ 58 pg (Ranjekar et al. 1978; Labani and Elkington 1987). Such contradictory data highlighted a need for clarification about the reliability of the nuclear DNA values published in the *Allium* genus.

Baranyi and Greilhuber (1999) reported that the 2C DNA values of the investigated *Allium* species were normally distributed, and only 29 of the 60 2C values published deviate less than 10% from the above results thus highlighting a need for improving a general agreement in the standardization and preparative procedures of cytophotometric genome size determination in *Allium* species.

The nuclear DNAs of two onion bacterial artificial chromosomes (BACs) were measured in order to study the existence and distribution of repetitive sequences as well as to assess gene densities (Jakše et al. 2008). The full sequences from two BACs were AT-rich (64.8%) and demonstrated lengthy tracts of degenerated retroviral elements and transposons, which were similar to those found in bigger plant

genomes. An estimated mean density of one gene per 168 kb was obtained from 499,997 bp of onion nuclear DNA, which is among the smallest reported to date. In comparison to random shotgun reads, methyl filtration reduced the frequency of anonymous sequences from 82% to 55% while increasing nonorganellar protein hits from 4% to 42%. (Jakše et al. 2008). Ricroch et al. (2005) evaluated the genome size of 30 *Allium* species including major vegetable crops and their wild allies. The 2C DNA value of gDNA was measured and compared with data derived from the angiosperms (http://www.rbgkew.org.uk). Overall, genome size (1C DNA) values varied from 7 pg in *A. altyncolicum* to, 31.5 pg in *A. ursinum*, showing 3.3-fold variation and a coefficient of variation of 27.4%. The GC content, evaluated in 24 accessions representing 23 *Allium* species, displayed values ranging from 38.5% in leek to 41.2% in *A. cernuum*, with a median value of 39.9% (Ricroch et al. 2005). Using different molecular markers, the comparison of molecular phylogenies led Fritsch and Friesen (2002) to propose a new idea for the classification of *Allium*, with 67 sections and 14 subgenera. A variation in cytogenetic characters such as nuclear DNA amount and chromosome numbers, when compared to the classification of the genus *Allium*, does not reveal clear discontinuities between the taxonomic groups (Ohri et al. 1998); but the variation mirrors the great diversity observed in morphology, life form, ecology, and breeding systems (Hanelt 1990).

4 Linkage, Cytogenetic, and Physical Mapping

The initial efforts have been succeeded by the construction of an *Allium* species genetic map using restriction fragment length polymorphism (RFLP) markers of the intraspecific onion cross 'BYG15-23 × AC43' (King et al. 1998). This map was subsequently extended with the SNP, and SSR markers, and homologs of transposable elements derived from the sequencing of expressed sequence tag (EST) in onion (Kuhl et al. 2004; Martin et al. 2005). These markers enabled partial map constructions in other intraspecific onion crosses to develop the map-based genetic analyses of various traits, including fertility restoration and bulb color development (Gokce et al. 2002; McCallum et al. 2006; Khar et al. 2008). More recently, another genetic map based on *A. fistulosum* genomic SSR markers and onion EST-derived SNP and SSR markers, providing further insight into comparative studies between the genomes of the two crops (Tsukazaki et al. 2010). The key resource that enabled alignment of *Allium* genetic maps to physical chromosomes and facilitated comparison between species was the complete set of *A. fistulosum-A. cepa* monosomic addition lines developed by Shigyo et al. (1996). These were initially applied to anchor AFLP-based maps in the interspecific *A. cepa*×*A. roylei* cross (van Heusden et al. 2000) and subsequently to anchor the 'BYG15-23 x AC43' map (Martin et al. 2005) and anchor SSR-based maps in *A. fistulosum* (Tsukazaki et al. 2008) to physical chromosomes and more recently to assign many more onion EST-derived anchor markers used in *A. fistulosum* maps (Tsukazaki et al. 2010). Further, a large number of molecular and phenotypic markers associated with many candidate genes coding for valuable economic traits, have also been assigned to A. *cepa* chromosomes (Martin et al. 2005; Masuzaki et al. 2006;

Yaguchi et al. 2008), providing an important guide for functional and QTL studies. Recently, RobustDb, a generic online genomics database that presented information mainly for garlic maps led to the development of marker data (Bhasi et al. 2010). Likewise, the VegMarks database (http://vegmarks.nivot.affrc.go.jp/) contained detailed information regarding *A. fistulosum* markers. A total of 107 markers comprising three *A. fistulosum* genomic SSRs, 31 SNP markers, and the 73 additional onion EST-SSRs derived from onion ESTs were evaluated in the population previously used to construct an AFLP-based linkage map (McCallum et al. 2012). Additionally, an interspecific *A. cepa* × *A. roylei* map was augmented with additional genetic markers to increase correspondences among *Allium* maps. Then, 11 linkage groups spanning one Morgan were formed using LOD 5 cutoff. Using the monosomic addition set, the anchor *loci* were assigned and mapped in the interspecific cross with the consequent additional landmarks for aligning genetic linkage maps in *A. cepa* and *A. fistulosum*. The AlliumMap database (http://alliumgenetics.org.) contains 512 correspondences between markers in various *Allium* maps (McCallum et al. 2012). This database represents an integrated point to access details of the genetic markers and sequence resources employed across multiple studies in cultivated *Allium* to resolve common questions of crop evolution and economic trait regulation across these major *Allium* crops. The genotyping-by-sequencing (GBS) provides a high degree of complexity reduction followed by synchronized SNP discovery and genotyping for species with complex genomes (Jo et al. 2017). GBS analysis conducted on F_2('NW-001' X 'NW-002') population, provided 1,851,428 SNPs. After filtering, 10,091 high-fidelity SNP markers that satisfied the criteria of segregation ratio were used to construct an onion genetic map (eight linkage groups) and spanned a genetic length of 1383 centiMorgans (cM) with an average marker interval of 8.08 cM. The newly developed linkage map will be valuable tools for the genetic mapping of important agronomic traits and marker-assisted selection in onion breeding programs worldwide. Likewise, the F_2[doubled haploid (DH) 'H6' (red bulb) and the inbred line 'SP3B' (yellow bulb)] population was used to construct another SNP-based genetic linkage map to identify the QTLs associated with the total anthocyanin content of onion bulbs using GBS analysis (Choi et al. 2020). Hence, an onion genetic map with eight linkage groups consisting of 319 GBS-based SNP loci and 34 high-resolution melting markers was constructed (Choi et al. 2020). Based on this linkage map, two QTLs, *qAC4.1* and *qAC4.2*, for anthocyanin content and one major QTL, *qAS7.1*, for anthocyanin synthesis were identified. This major QTL located on chromosome 7 with a phenotypic variation of 87.61% can be putatively identified as *dihydroflavonol 4-reductase* (*DFR*) gene responsible for bulb coloring in onion. Recently, an ultra-high-density linkage map of transcriptome-based unigene markers was assigned on *A. cepa* chromosomes using RNA sequence data from 96 lines of the F_2 mapping population (a doubled haploid shallot 'DHA' × a doubled haploid onion 'DHC') (Fujito et al. 2021). SNPs with at least four reads on 25,462 unigenes were anchored on eight *A. cepa* chromosomes. In total, multiple SNPs were identified on 22,184 unigenes and a single SNP site on 3278. The chromosome marker information was made public via the web database *Allium* TDB (http://alliumtdb.kazusa.or.jp/), which is valuable in genome sequencing and useful for validation of the original genetic map in *Allium.* In

a recent study, an ultrasensitive dual-color Tyr-FISH technique was developed to construct a reliable and easy method for in situ mappings of short unique DNA sequences on onion chromosomes (Kudryavtseva et al. 2021), which allows more accurate determination for physical distances between markers and can be applied for the faster integration of genetic and cytogenetic maps. For instance, Tyr-FISH was applied to physically locate molecular markers tightly linked to the nuclear male-fertility (Ms) restoration locus on onion (Khrustaleva et al. 2016). Relatively short genomic amplicons and a cDNA clone were successfully visualized on chromosome 2 near the centromere, a region of low recombination, which partially explained the identified molecular markers tightly linked to the Ms. locus in different segregating progenies (Khrustaleva et al. 2016). Likewise, Tyr-FISH mapping on *A. cepa* chromosomes was performed using EST-clones API20 and API66, previously mapped in the linkage groups corresponding to *A. cepa* chromosomes1 and 5, respectively (Romanov et al. 2014). Results indicated that API66 possesses insert with high homology to the sucrose transportase gene, and API20 – to the light harvesting-like protein 3 encoding gene (Romanov et al. 2014). In addition, Tyr-FISH analysis of EST-clone API66 revealed 25 hybridization sites on the mitotic metaphase chromosomes of *A. cepa*, and the signals were dispersed along all chromosomes (Romanov et al. 2014). It is worth noting that the sucrose transportase gene belongs to a large superfamily of genes in plants, which explained multiple sites of API66 hybridization (Romanov et al. 2014). Future research involves the creation of more molecular cytogenetic markers based on mapping chromosome-specific repeats and genes/markers can be further applied in onion breeding and genome assembly at the chromosome level. Expansion of FISH applications in plants with the development of new technologies as oligo-FISH mapping, CAS-FISH for living cells, RNA-FISH for measuring gene transcription, and a haplotype-specific FISH for crossing over visualization will be gold stand method for mapping physical chromosomes and detection their organization.

5 Novel Analysis Methods for Large Genome Size

Irrespective of ploidy, genome sizes refer to the amount of DNA enclosed in a haploid genome set expressed in terms of the number of base pairs (1 kb = 1000 bp), megabases (1 Mb = 1,000,000 bp), the DNA mass in picograms (1 pg = 10^{-12} g = 1000 Mb of DNA), or as C-value ('C' stands for constant referring to the fact that the genome size is constant from cell to cell in a given organism), equals to the mass of DNA in Pg) in a haploid set of chromosomes (Maloy and Hughes 2013), regardless of ploidy level which means one-third or one-fourth of the total mass of DNA in the cell (Maloy and Hughes 2013).

The genome size depends on differences in the rates at which insertions and deletions (indels) occur, transposon activity, segmental duplications, and the efficiency of natural selection in eliminating or promoting such changes (Sung et al. 2016). Large genomes have higher mutational liability, even in noncoding areas,

thus, a large fraction of noncoding regions is likely to evolve neutrally concerning indels (Mueller 2015).

Bulb onion has a large haploid genome size of 16,400 Mb and unique telomeric sequences with the lowest GC content (32%) among angiosperms (Abdelrahman et al. 2020a, b; Shigyo et al. 2018), yet it is the largest of all diploid crops (Brenchley et al. 2012; Marcussen et al. 2014). Enormous genome-size differences exist among closely related *Allium* species. For example, bulb onion has 28% more nuclear DNA than *A. fistulosum* and 7% more than garlic nuclear DNA.

The bulb onion genomic fragments cloned into BACs demonstrated that most sequences were similar to retro-element and transposons, indicating that the onion genome exhibited low gene densities and high frequencies of repetitive DNA as major components of large chromosomes (Suzuki et al. 2001). The C_oT reannealing kinetics indicated that about 40% of the onion genome is highly repetitive ($>1000\times$ copies) and 40% has 100–1000 copies (Stack and Comings 1979). Overall, at least 95% of the *A. cepa* genome consists of repetitive sequences, most of which are dispersed repeats (Shibata and Hizume 2002). Due to the size of the genome and its repetitive nature, developing an onion reference genome assembly is challenging.

Although the accessibility of the structural genomic resources in *Allium* species is limited, using next-generation sequencing (NGS; e.g., Roche, Illumina, and ABI Solid platforms) technologies and large computational abilities, sequencing, assembly, and annotation of large genomes has become feasible. The advantages of third-generation sequencing (TGS) technologies such as Pacific Biosciences over short-read sequencing as Illumina's HiSeq, NovaSeq, NextSeq, MiSeq instruments, and Ion Torrent sequencers (Thermo Fisher Scientific) include greater read lengths, thus facilitating the assembly of large and complex genomes such as onion (Abdelrahman et al. 2018a, b), i.e., reads of up to 600 bases (Goodwin et al. 2016), and >10 kb (Rothberg et al. 2011), respectively. Consequently, improved *de novo* assembly, mapping certainty, transcript isoform identification, and detection of structural variants become feasible (Depledge et al. 2019). Concomitantly, the ever-diminishing computational power costs and the growing availability of on-demand cloud-based computing steadily increase assembly and annotation efficiency. Hence, large-scale and cost-effective sequencing, assembly, and annotation of onion DNA are becoming feasible (Abdelrahman et al. 2020a, b).

The three key approaches to large genome sequencing like the onion genome include the shot-gun sequencing, where DNA is fragmented into short reads, which are then sequenced. Since a large proportion of onion reads are repeated sequences, shot-gun sequencing is unsuitable for its large and complex genome. Suzuki et al. (2001) were the first to Sanger sequence large genomic fragments cloned into BACs. Two entire BAC clones, and the random ends of genomic fragments cloned into BACs, yielded 298 kb of random-end sequences and 202 kb from the entire BAC clones (Jakše et al. 2008), with AT content of 35.7%, which is close to the estimate of 32% by Kirk et al. (1970). Additionally, only 5% of these random genomic sequences showed significant similarities to nonorganellar proteins, 25% were highly similar to transposons or retrotransposons, and 70% showed no significant hits to the databases and were primarily degenerated retroelements (Jakše et al.

2008). Additional sequence survey of onion random genomic sequences was conducted using DNA isolated from DH line 15,197. The DNA was trimmed, and random fragments were sequenced using Sanger and 454 technologies, generating 6590 reads, among them, 82% had no significant hits in the databases, 4% matched nuclear-encoded proteins, and 14% matched transposons (Jakše et al. 2008).

The second approach efficiently samples expressed regions. Sequencing random cDNAs thus revealed the collection of genes (the transcriptome) expressed in any given tissue, at a specific developmental stage, or after treatment (Abdelrahman et al. 2017a, b, c, 2020a, b). However, the normalization of the cDNA reduces the frequencies of highly expressed genes and increases the number of sequencing reads from rarer transcripts (Abdelrahman et al. 2017a, b, c).

A free access web-based tool named *Allium* TDB (http://alliumtdb.kazusa.or.jp/) contains several transcriptional data of *Allium* species. Likewise, the transcriptome sequencing on vegetative and reproductive tissues of homozygous DH CUDH2107 was used to develop a multi-organ reference transcriptome catalogue (Khosa et al. 2016). In general, the 271,665 contigs of transcripts were generated using the Trinity pipeline. This dataset was analyzed for gene ontology (GO), and the bulb onion transcripts were grouped into 95 functional groups, and among which Binding Domains the most abundant GO term followed by Carbohydrate-Binding (Khosa et al. 2016).

The third approach involves enrichment for lower-copy regions by removing repetitive DNAs, and reduced representation sequencing involves selection against methylated DNA (Rabinowicz et al. 1999). In bulb onion, the sequencing of these fragments is an efficient approach to enrich genic regions in its enormous genome. Hence, the complete onion DH 15197 sequencing of methyl-filtered DNA fragments revealed that among the 2712 methyl-filtered fragments, 3% matched transposons, 55% were anonymous, and 42% were similar to nonorganellar proteins (Jakše et al. 2008).

The first insight into the genome sequencing of *A. cepa* is now ongoing through the SEQUON project (http://www.oniongenome.net) using Illumina HiSeq 2500, intending to sequence the onion nuclear DNA and provide de novo genome assembly and ab initio annotation thus applied to the doubled haploid onion line DHCU066619, for a whole-genome shotgun sequencing of an onion genome (Finkers et al. 2021). Initially, Illumina HiSeq 2500 sequencing of 3 small libraries yielded 769 Gb sequence data, whose analyses provided an estimated genome size of ~13.6 Gb. Secondly, the MaSurCa based assembly resulted in 10.8 Gb in 6.2 M contigs with a contig N50 of 2.7Kb. This assembly was further scaffolded using 18.1 M PacBio RS II reads. Finally, the hybrid Illumina/PacBio assembly was further improved using Dovetail Chicago and subsequent HiRise scaffolding, thus indicating no misassemblies in the Illumina/PacBio hybrid contigs. The combination of these 3 technologies resulted in an assembly of 14.9 Gb in 92.9 K scaffolds with a scaffold N50 size of 436Kb. With an estimated genome size of 164 Gb/1C. To further organize genome assembly, three intraspecific genetic linkage maps (Duangjit et al. 2013; Fujito et al. 2021; Choi et al. 2020) and two interspecific genetic linkage maps (Scholten et al. 2016) were used to anchor scaffolds into pseudomolecules using AllMaps (Tang et al. 2015). A final assembly of 14.9 Gb with an N50 of 461 Kb. out of which, only 2.2 Gb (~6.7%) was ordered into

8 pseudomolecules using 5 different genetic linkage maps (Finkers et al. 2021), whereas the remaining of the genome is presented in 89.8 K scaffolds. Additionally, 72.4% of the genome was successfully identified as repetitive sequences consisting mainly of retrotransposons. The ab initio gene prediction indicated 540,925 putative gene models, which is far more than expected, possibly due to the presence of pseudogenes. Of these models, 86,073 showed similarity to UNIPROT published proteins. No gene-rich regions were found, and most genes were uniformly distributed across the genome (Finkers et al. 2021). The analysis of synteny with garlic showed collinearity but also major rearrangements between both species. This assembly is the first high-quality genome sequence available for the study of onion and will be a valuable resource for further research.

Similarly, garlic chromosome-level reference genomes and annotations by combining SMRT sequencing with PacBio Sequel Nanopore, 10x Genomics, Illumina HiSeq paired-end, and high-throughput chromosome conformation capture (Hi-C) sequencing resulted in a genome of 16.24 Gb (contig N50 length = 194 kb, scaffold N50 length = 725 kb (Sun et al. 2020). Then, 252.5-Gb of long reads obtained from Nanopore sequencing were mapped to the assembly. Thereafter, the Hi-C technology was applied to assemble a chromosome-level genome. A total of 6.34 billion reads from four Hi-C libraries were used, and approximately 87.5% of the assembled sequences were mapped to 8 pseudomolecules with a super-scaffold N50 length of 1.69 Gb (Sun et al. 2020). The final assembly covered 96.1% of the garlic genome (approximately 16.9 Gb), as estimated by the k-mer analysis. In addition, a total of 38 alliinase genes were expressed in various tissues, out of which, 4 tandem duplicated genes (*Asa5G05557.1*, *Asa5G05559.1*, *Asa5G05560.*1, and *Asa5G05561.1*) produced highly abundant transcripts in bulb tissues (Sun et al. 2020). In addition, their expression levels increased during bulb growth and expansion, suggesting that these four genes exhibited functional redundancy and may be responsible for the biosynthesis of allicin at a high concentration in garlic bulbs (Sun et al. 2020).

6 Metabolomic and Transcriptomic Landscapes of *Allium* Crops in Response to Environmental Stress

Because of the high yield losses in principle crops caused by global warming in addition to the rapidly increasing consumption of food, there is an urgent need to boost food security (Abiala et al. 2018; Zhang et al. 2019). Nevertheless, there have been few studies that address the Allium metabolome and transcriptome profiling in the context of environmental stress, so the Allium international community may need to make more efforts in this area. Transcriptome analysis of commercial brown onion cv. 'Orlando' inner and outer scales in response to heat stress revealed that oxidation and lipid metabolism pathways, in addition to cell-wall adjustment, were abundantly expressed in the onion outer scale under heat stress (Galsurker et al. 2018). However, defense response-related genes, such as those gene that encodes antioxidant activities, heat shock proteins, or even the manufacturing of osmoprotectant metabolites, were highly stimulated in the inner scale. (Galsurker et al. 2018). These transcriptomic data

led to the creation of a conceptual framework that proposes sequential processes for the advancement of desiccation and browning on the outer scale versus processes associated with defense response and high temperature tolerance in the inner scales (Galsurker et al. 2018). Transcriptome sequencing of cold-tolerant and cold-sensitive onion genotypes under freezing and cold conditions revealed that freezing and cold stresses significantly induced numerous genes in tolerant lines compared to susceptible genotypes (Han et al. 2016). Among these transcript, genes-encoding hypothetical proteins, heat shock proteins (HSPs), zinc finger (ZIP) proteins, and CBL-interacting protein kinase (CIPK) , as well as a subset of transcription factors (TFs), especially TFs that function as activators including dehydration-responsive element (DRE)-binding (DREB), CBL, MYB, bZIP, zinc finger of *Arabidopsis thaliana* (ZAT), HSPs, and basic helix-loop-helix (bHLH) were radically changed during freezing and cold stress conditions (Han et al. 2016). Likewise, a genome-wide transcriptome profiling analysis of garlic under cold stress revealed that enzyme-encoding genes, which knowingly enriched in the "proteasome" pathway such as γ-glutamyltranspeptidase-, δ-aminolevulinic acid dehydratase-, and alliinase-encoding genes are conceivably implicated in garlic discoloration under low temperature stress (Li et al. 2018). These stress-responsive genes could be to blame for garlic discoloration caused by low temperatures (Li et al. 2018). Because environmental stress impacts plant growth, identifying stress biomarkers is an important prerequisite for breeding stress-tolerant crops. In this respect, shallots are recognized as an important genetic resource for the breeding of common onions due to their high versatility to subtropical and tropical climates (Abdelrahman et al. 2015). The bulb onion double haploid, shallot double haploid, and its F1 hybrid were checked using LC-QqQ-MS. There were 113 targeted metabolites found in total, and the principal component analysis and volcano plot analysis clearly revealed genotype-specific metabolites that can be used as metabolic markers of environmental tolerance (Abdelrahman et al. 2015). Likewise, incorporated transcriptome and metabolome analysis of *A. fistulosum* with extra shallot chromosome 5A disclosed an accumulation of many flavonoids that are important in abiotic and biotic stress tolerance (Abdelrahman et al. 2019). The increase in flavonoid pool in A. fistulosum with extra chromosome 5A from shallot was also coherent with the increased expression of many upstream and downstream flavonoid biosynthesis and regulatory genes (Abdelrahman et al. 2019). The above study proved that shallot can be used as a genetic resource to enhance onion stress tolerance. Similarly, Zhang et al. (2018) used transcriptome analysis of two contrasting dark-red and white onion cultivars to discover that both *flavonoid 3′,5′-hydroxylase* (*F3′,5′H*) and *dihydroflavonol 4-reductase* (*DFR*) genes play important roles in the biosynthesis of dark-red bulbs, and that flavonol synthase (*FLS*) and *DFR gene* expression levels could respond to prevent blue coloration. In addition, the positive variation in the F3′,5′H/F3′H ratio also affects onion bulb color diversity (Zhang et al. 2018). The metabolic characteristics in the bulbs of eight Indonesian shallot landraces and seven short-day and three long-day bulb onion cultivars have been identified using LC-Q-TOF-MS/MS in order to generate new genetic materials for the development of a novel bulb onion cultivar derived from intraspecific hybrids with beneficial agronomic traits from shallots (Abdelrahman et al. 2020a, b). The results showed that free and

conjugated amino acids, flavonoids (particularly metabolites containing flavonol aglycone), anthocyanins, and organic acids were among the top metabolite factors that were largely correlated with shallot landraces when compared to bulb onion cultivars (Abdelrahman et al. 2020a, b). Furthermore, the measurement of 21 amino acids using traditional HPLC analysis revealed that shallots had elevated concentrations than bulb onions (Abdelrahman et al. 2020a, b). The current study found that shallots reprogrammed their metabolism to accumulate more amino acids and flavonoids as an adaptation to incredibly hot tropical environments.

7 Prospects

Omics' techniques such as transcriptomics, proteomics, metagenomics, metabolomics, and genomics have the potential to provide new avenues of research as well as advancement for Allium crops with restricted genome sequence knowledge. NGS systems have lately been regularly used to reveal markers such as SSRs and SNPs in numerous bulb onion populations. For example, a high-throughput genotyping method, such as diversity arrays (DArTseq), was used to analyze population dynamics in a large garlic population with 417 accessions. (Egea et al. 2017). Likewise, SNPs were discovered using the double digest restriction site-associated DNA sequencing (ddRAD-seq) method in inbred lines of Korean short-day onion, allowing for structure analyses and genetic relationship studies (Lee et al. 2018). Even though genomic resources in Allium species have been limited to date, the latest rapid advancement of NGS technology has rendered whole-genome sequencing much easier than previously. The genome sequencing of *A. cepa* is now ongoing through the SEQUON project (http://www.oniongenome.net) by using Illumina HiSeq 2500 sequencer. More recently, HiFi reads [a type of data produced using the circular consensus sequencing (CCS) mode on one of the PacBio Sequel Systems], provide base-level resolution with >99.9% single-molecule read accuracy. HiFi reads can accurately detect all types of variants, from single nucleotide to structural variants, with high precision and recall and phase haplotypes, even in hard-to-sequence regions of complex genomes such as onion.

Transcriptome sequencing is an efficient approach to generate sequence information from expressed regions of the genome and has been widely used in the Alliums. Several workers use RNA-seq for studying organ development, marker discovery, male sterility, flavonoid biosynthesis, the abiotic stress response in different *Allium* species (Abdelrahman et al. 2015; Abdelrahman et al. 2019). In context, *Allium*TDB (http://alliumtdb.kazusa.or.jp/) contains many useful transcript libraries of root, stem, bulb, and leaves for various *Allium* species. Similarly, a draft reference transcript for onion, using long-read sequencing technology will help in whole-genome sequencing in Alliums (Sohn et al. 2016). A transcriptome catalogue for homozygous double haploid line 'CUDH2107' for understanding the genetic and molecular basis of various traits was developed (Khosa et al. 2016).

After the establishment of omics platforms, plant metabolism research has transitioned from the study of individual gene functions to metabolic system research.

Mass spectrometry (MS) is a major instrument that has high sensitivity and broad metabolite detection capabilities, thus facilitated the detection of hundreds of metabolites having diverse bioactivities in the alliaceous crops (Abdelrahman et al. 2020a, b, 2021a, b). In the future, the individual bio-resource-specific metabolic patterns can be used for molecular breeding of *Allium* crops while the broad metabolic profiles of *Allium* bio-resources can be used for integrated omics approaches. The integration of metabolomics and transcriptomics will provide insight into the molecular mechanism of *Allium* metabolite biosynthesis.

Image techniques such as GISH and FISH using fluorescent imaging systems are important approaches for *Allium* breeding and genetic studies, thus further developments in *Allium* imaging analysis are needed. Recently, the selection of fusarium basel rot-resistant onion was carried out using digital image analysis to quantify symptom development in the basal plate of dormant bulbs (Mandal and Cramer 2021). Analysis with confocal microscopy identified bright blue-green autofluorescence from *Fusarium oxysporum*-infected tissue, effectively differentiating diseased from healthy tissue. Visual scoring of the fusarium basel rot symptom was combined by stereo fluorescence microscopic images, captured using a green fluorescence protein dual filter to quantify accurately fusarium basel rot severity in the basal plate tissue (Mandal and Cramer 2021). The new developed method could be used for developing resistant cultivars for onion breeding programs in the near future.

Targeted genome-editing technologies, especially clustered regularly interspaced short palindromic repeats (CRISPR)/(CRISPR)-associated protein 9 (Cas9), have great potential to aid in the breeding of crops that can produce high yields under biotic/abiotic stress (Abdelrahman et al. 2018a, b). This is due to their high accuracy and low risk of off-target effects, compared with conventional methods. The use of the CRISPR/Cas9 system is commonly used for targeted mutagenesis in crop plants, including gene knockouts, modifications, and the activation and repression of target genes. However, there are no reports regarding genome editing in onion probably due to the complex genome structure with many repetitive sequences, affecting the efficacy of the CRISPR/Cas9 system. Genomic selection (GS) is being increasingly applied in plant breeding programs to enhance the genetic gain of economically important traits. For example, the GS scheme was proposed to prevent inbreeding depression in onions by avoiding the co-selection of closely related plants and combining the shortening of generation time (Sekine and Yabe 2020). Ten years of breeding programs to evaluate the efficiency of different selection schemes in onion, including general phenotypic selection, self-crossing phenotypic selection, general genomic selection, and inbreeding avoiding genomic selection were compared (Sekine and Yabe 2020). General GS with shortening of generation time yielded the highest genetic gains among the selection schemes, however, inbreeding depression increased rapidly in later years. The proposed GS combining shortening of generation time with updating of the prediction model was superior to the others in later years, as it yielded relatively high genetic gain while maintaining significantly low levels of inbreeding. These results suggested that GS can be beneficial in onion breeding, and an optimal scheme should be selected depending on the selection period.

8 Conclusion

In conclusion, the study of metabolome and cytoplasmic genomics of *Allium* holds significant promise in advancing our understanding of this plant genus. By analyzing the metabolites present in the cytoplasm and investigating the genetic information within the organelles, researchers can gain valuable insights into the biochemical pathways and genetic mechanisms that underlie various traits and processes in *Allium* species. This knowledge could have broad implications for agriculture, medicine, and biotechnology, potentially leading to improved crop yields, novel pharmaceutical compounds, and more sustainable farming practices. However, further research and collaboration are needed to fully unravel the complex interplay between metabolome and cytoplasmic genomics in *Allium* and unlock its full potential for practical applications.

References

Abdelrahman M (2020) *Allium* functional genomic development for future climatic changes. In: Kole C (ed) Genomic designing of climate-smart vegetable crops. Springer, Cham, pp 345–358

Abdelrahman M, Hirata S, Ito S-I, Yamauchi N, Shigyo M (2014) Compartmentation and localization of bioactive metabolites in different organs of *Allium roylei*. Biosci Biotechnol Biochem 78:1112–1122

Abdelrahman M, Sawada Y, Nakabayashi R, Sato S, Hirakawa H et al (2015) Integrating transcriptome and target metabolome variability in doubled haploids of *Allium cepa* for abiotic stress protection. Mol Breed 35:195

Abdelrahman M, Abdel-Motaal F, El-Sayed M, Jogaiah S, Shigyo M, Ito S-I, Tran LS (2016) Dissection of *Trichoderma longibrachiatum*-induced defense in onion (*Allium cepa* L.) against *Fusarium oxysporum* f. sp. *cepa* by target metabolite profiling. Plant Sci 246:128–138

Abdelrahman M, Suzumura N, Mitoma M, Matsuo S, Ikeuchi T, Mori M et al (2017a) Comparative *de novo* transcriptome profiles in *Asparagus officinalis* and *A. kiusianus* during the early stage of *Phomopsis asparagi* infection. Sci Rep 7:2608

Abdelrahman M, El-Sayed M, Sato S, Hirakawa H, Ito SI et al (2017b) RNA-sequencing-based transcriptome and biochemical analyses of steroidal saponin pathway in a complete set of *Allium fistulosum-A. cepa* monosomic addition lines. PLoS One 12:0181784

Abdelrahman M, Mahmood HYAH, El-Sayed M, Tanaka S, Tran LS (2017c) Isolation and characterization of Cepa2, a natural alliospiroside A, from shallot (*Allium cepa* L. Aggregatum group) with anticancer activity. Plant Physiol Biochem 116:167–173

Abdelrahman M, Al-Sadi AM, Pour-Aboughadareh A, Burritt DJ, Tran LP (2018a) Genome editing using CRISPR/Cas9–targeted mutagenesis: an opportunity for yield improvements of crop plants grown under environmental stresses. Plant Physiol Biochem 131:31–36

Abdelrahman M, Jogaiah S, Burritt DJ, Tran LP (2018b) Legume genetic resources and transcriptome dynamics under abiotic stress conditions. Plant Cell Environ 41:1972–1983

Abdelrahman M, Hirata S, Sawada Y, Hirai MY, Sato S et al (2019) Widely targeted metabolome and transcriptome landscapes of *Allium fistulosum–A. cepa* chromosome addition lines revealed a flavonoid hot spoton chromosome 5A. Sci Rep 9:3541

Abdelrahman M, Nakabayashi R, Mori T, Ikeuchi T, Mori M, Murakami K et al (2020a) Comparative metabolome and transcriptome analyses of susceptible *Asparagus officinalis* and resistant wild *A. kiusianus*revealinsights into stem blight. Plant Cell Physiol 61:1464–1476

Abdelrahman M, Ariyanti NA, Sawada Y, Tsuji F, Hirata S, Hang TTM et al (2020b) Metabolome-based discrimination analysis of shallot landraces and bulb onion cultivars associated with differences in the amino acid and flavonoid profiles. Molecules 25:5300

Abdelrahman M, Nishiyama R, Tran CD, Kusano M, Nakabayashi R et al (2021a) Defective cytokinin signaling reprograms lipid and flavonoid gene-to-metabolite networks to mitigate high salinity in *Arabidopsis*. Proc Natl Acad Sci U S A 118(48):e2105021118

Abdelrahman M, Hirata S, Mukae T, Yamada T, Sawada Y et al (2021b) Comprehensive metabolite profiling in genetic resources of garlic (*Allium sativum* L.) collected from different geographical regions. Molecules 26:1415

Abiala MA, Abdelrahman M, Burritt DJ, Tran LP (2018) Salt stress tolerance mechanisms and potential applications of legumes for sustainable reclamation of salt-degraded soils. Land Degrad Dev 29:3812–3822

Alan AR, Toprak FC, Kaska A (2016) Production and evaluation of gynogenic leek (*Allium ampeloprasum* L.) plants. Plant Cell Tissue Organ Cult 125:249–259

Albini SM, Jones GH (1990) Synaptonemal complex spreading in *Allium cepa* and *Allium fistulosum*. III. The F_1 hybrid. Genome 33:854–866

Al-Sheikh Hussain L (1977) Cytology and species relationships of *Allium* section Codonoprasum (Rchb.) Endl. PhD thesis, University of Sheffield

Ariyanti NA, Hoa VQ, Khrustaleva LI, Hirata S, Abdelrahman M et al (2015) Production and characterization of alien chromosome addition lines in *Allium fistulosum* carrying extra chromosomes of *Allium roylei* using molecular and cytogenetic analyses. Euphytica 206:343–355

Baranyi M, Greilhuber J (1999) Genome size in *Allium*: in quest of reproducible data. Ann Bot 83: 687–695

Bennett MD, Bhandol P, Leitch IJ (2000) Nuclear DNA amounts in angiosperms and their modern uses – 807 new estimates. Ann Bot 86:859–909

Bhasi A, Simon P, Senalik D, Kumar B, Manikandan V et al (2010) RoBuST: an integrated resource of genomics information for plants in the root and bulb crop families Apiaceae and Alliaceae. BMC Plant Biol 10:161

Bohanec B (2002) Double-haploid onion. In: Rabinowitch HD, Currah L (eds) *Allium* crop science: recent advances. CABI, Wallingford, pp 145–157

Bohanec B, Jakše M (1999) Variations in gynogenic response among long-day onion (*Allium cepa* L.) accessions. Plant Cell Rep 18:737–742

Brenchley R, Spannagl M, Pfeifer M, Barker GL, D'Amore R et al (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491(418):705–710

Cho KS, Yun BK, Yoon YH, Hong SY, Mekapogu M, Kim KH, Yang TJ (2015) Complete chloroplast genome sequence of tartary buckwheat (*Fagopyrum tataricum*) and comparative analysis with common buckwheat (*F esculentum*). PLoS One 10:e0125332

Choi Y, Kim S, Lee J (2020) Construction of an onion (*Allium cepa* L) genetic linkage map using genotyping-by-sequencing analysis with a reference gene set and identification of QTLs controlling anthocyanin synthesis and content. Plan Theory 9:616

De Melo PE (2003) The root systems of onion and *Allium fistulosum* in the context of organic farming: a breeding approach. PhD thesis, Wageningen University

Depledge DP, Srinivas KP, Sadaoka T, Bready D, Mori Y, Placantonakis DG et al (2019) Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. Nat Commun 10:754

Duangjit J, Bohanec B, Chan AP, Town CD, Havey MJ (2013) Transcriptome sequencing to produce SNP-based genetic maps of onion. Theor Appl Genet 126:2093–2101

Duangjit J, Welsh K, Wise ML, Bohanec B, Havey MJ (2014) Genetic analyses of anthocyanin concentrations and intensity of red bulb color among segregating haploid progenies of onion. Mol Breed 34:75–85

Egea LA, Mérida-García R, Kilian A, Hernandez P, Dorado G (2017) Assessment of genetic diversity and structure of large garlic (*Allium sativum*) germplasm bank, by diversity arrays technology genotyping-by-sequencing platform (DArTseq). Front Genet 8:98

Emsweller SL, Jones HA (1935) An interspecific hybrid in *Allium*. Hilgardia 9:265–273

Fayos O, Vallés MP, Garcés-Claver A, Mallor C, Castillo AM (2015) Doubled haploid production from Spanish onion (*Allium cepa* L) germplasm: embryogenesis induction, plant regeneration and chromosome doubling. Front Plant Sci 6:384

Finkers R, van Kaauwen M, Ament K, Burger-Meijer K, Egging R, Huits H et al (2021) Insights from the first genome assembly of onion (*Allium cepa*). bioRvix 0305434149

Fritsch RM, Friesen N (2002) Evolution, domestication, and taxonomy. In: H. D. Rabinowitch & L. Currah (Eds.), Allium Crop Science: Recent Advances, pp. 5–30. CABI Publishing, Wallingford, Oxon OX10 8DE, UK.

Fritsch RM, Blattner FR, Gurushidze M (2010) New classification of *Allium* L.subgMelanocrommyum (Webb & Berthel) Rouy (Alliaceae) based on molecular and morphological characters. Phyton 49:145e220

Fujito S, Akyol TY, Mukae T, Wako T, Yamashita K-I, Tsukazaki H, Hirakawa H et al (2021) Construction of a high-density linkage map and graphical representation of the arrangement of transcriptome-based unigene markers on the chromosomes of onion, *Allium cepa* L. BMC Genomics 22:481

Galsurker O, Doron-Faigenboim A, Teper-Bamnolker P, Daus A, Lers A, Eshel D (2018) Differential response to heat stress in outer and inner onion bulb scales. J Exp Bot 69:4047–4064

Galván Vivero GA (2009) Resistance to Fusarium basal rot and response to arbuscular mycorrhizal fungi in *Allium*. PhD thesis, Wageningen University

Galván Vivero GA, Kuyper TW, Burger K, Paul Keizer LC, Hoekstra RF, Kik C, Scholten QE (2011) Genetic analysis of the interaction between *Allium* species and arbuscular mycorrhizal fungi. Theor Appl Genet 122:947–960

Gokce A, McCallum J, Sato Y, Havey M (2002) Molecular tagging of the *Ms* locus in onion. Am Soc Hortic Sci 127:576–582

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17:333–351

Govaerts R, Kington S, Friesen N, Fritsch RM, Snijman DA et al (2005–2020) World checklist of Amaryllidaceae. https://apps.kew.org/wcsp/. Accessed 30 Oct 2022

Han J, Thamilarasan SK, Natarajan S, Park J-I, Chung M-Y, Nou I-S (2016) *De Novo* assembly and transcriptome analysis of bulb onion (*Allium cepa* L) during cold acclimation using contrasting genotypes. PLoS One 11:e0161987

Hanelt P (1990) Taxonomy, evolution, and history. In: H. D. Rabinowitch & J. L. Brewster (Eds.), Onions and Allied Crops. Vol. I, pp. 1–26. CRC Press, Boca Raton, Florida

Hang TTM, Shigyo M, Yaguchi S, Yamauchi N, Tashiro Y (2004) Effect of single alien chromosome from shallot (Allium cepa L. Aggregatum group) on carbohydrate production in leaf blade of bunching onion (A. fistulosum L.). Genes Genet Syst 79:345–350

Hao M, Zhang L, Ning S, Huang L, Yuan Z et al (2020) The resurgence of introgression breeding, as exemplified in wheat improvement. Front Plant Sci 11:252

Havey MJ (2002) Genome organization in Allium 59–79. Rabinowitch HD, Currah L Allium crop science: Recent advances CABI Wallingford, UK

Hu H, Hu Q, Al-Shehbaz IA, Luo X, Zeng T, Guo X, Liu J (2016) Species delimitation and interspecific relationships of the genus *Orychophragmus* (Brassicaceae) inferred from whole chloroplast genomes. Front Plant Sci 7:1826

Hui L, Chen W, Pan M, Li W, He L, Wang J, Miao M, Yang H (2020) Characterization of the complete mitochondrial genome of yellow stripe mutant in onion (*Allium cepa* L). Mitochondrial DNA B 5:2986–2987

Huo YM, Gao LM, Liu BJ, Yang YY, Kong SP, Sun YQ, Yang YH, Wu X (2019) Complete chloroplast genome sequences of four *Allium* species: comparative and phylogenetic analyses. Sci Rep 9:12250

Hyde PT, Earle ED, Mutschler MA (2012) Doubled haploid onion (*Allium cepa* L) lines and their impact on hybrid performance. Hortic Sci 47:1690–1695

Ibrahim AM, Kayat FB, Susanto D, Kashiani P, Arifullah M (2016) Haploid induction in spring onion (*Allium fistulosum* L) via gynogenesis. Biotechnology 15:10–16

Jakše J, Meyer JDF, Suzuki G, McCallum J, Cheung F, Town CD, Havey MJ (2008) Pilot sequencing of onion genomic DNA reveals fragmented transposable elements, low gene densities, and significant gene enrichment after methyl filtration. Mol Genet Genom 280:287–292

Jo J, Purushotham PM, Han K, Lee H-R, Nah G, Kang B-C (2017) Development of a genetic map for onion (*Allium cepa* L) using reference-free genotyping-by-sequencing and SNP assays. Front Plant Sci 8:1606

Kaska A, Toprak FC, Alan AR (2013) Gynogenesis induction in leek (*Allium ampeloprasum* L) breeding materials. Curr Opin Biotechnol 24S:S42–S47

Kato N, Reynolds D, Brown ML, Boisdore M, Fujikawa Y, Morales A, Meisel LA (2008) Multidimensional fluorescence microscopy of multiple organelles in *Arabidopsis* seedlings. Plant Methods 4:9–10

Keller J (1990) Culture of unpollinated ovules, ovaries, and flower buds in some species of the genus *Allium* and haploid induction via gynogenesis in onion (*Allium cepa* L). Euphytica 47:241–247

Kelly LJ, Leitch IJ (2011) Exploring giant plant genomes with next-generation sequencing technology. Chromosom Res 19:939–953

Khan PSSV, Vijayalakshmi G, Raja MM, Naik ML, Germanà MA, Terry RG (2020) Doubled haploid production in onion (*Allium cepa* L): from gynogenesis to chromosome doubling. Plant Cell Tissue Organ Cult 142:1–22

Khar A, Jakse J, Havey M (2008) Segregations for onion-bulb colors reveal that red is controlled by at least three loci. Am Soc Hortic Sci 133:42–47

Khosa JS, Lee R, Bräuning S, Lord J, Pither-Joyce M, McCallum J, Macknight RC (2016) Doubled Haploid 'CUDH2107' as a Reference for bulb onion (*Allium cepa* L) research: development of a transcriptome catalogue and identification of transcripts associated with male fertility. PLoS One 11:e0166568

Khrustaleva LI, Kik C (1998) Cytogenetical studies in the bridge cross *Allium cepa* × (*A fistulosum* × *A roylei*). Theor Appl Genet 96:8–14

Khrustaleva LI, de Melo PE, van Heusden AW, Kik C (2005) The integration of recombination and physical maps in a large-genome monocot using haploid genome analysis in a trihybrid *Allium* population. Genetics 169:1673–1685

Khrustaleva L, Jiang J, Havey MJ (2016) High-resolution tyramide-FISH mapping of markers tightly linked to the male-fertility restoration (Ms) locus of onion. Theor Appl Genet 129: 535–545

Kim B, Kim K, Yang T-J, Kim S (2016) Completion of the mitochondrial genome sequence of onion (*Allium cepa* L) containing the CMS-S male-sterile cytoplasm and identification of an independent event of the ccmF N gene split. Curr Genet 62:873–885

Kim B, Yang T-J, Kim S (2018) Identification of a gene responsible for cytoplasmic male-sterility in onions (*Allium cepa* L) using comparative analysis of mitochondrial genome sequences of two recently diverged cytoplasms. Theor Appl Genet 132:313–322

King J, Bradeen J, Bark O, McCallum J, Havey M (1998) A low-density genetic map of onion reveals a role for tandem duplication in the evolution of an extremely large diploid genome. Theor Appl Genet 96:52–62

Kirk JTO, Rees H, Evans G (1970) Base composition of nuclear DNA with the genus *Allium*. Heredity 25:507–512

Kudryavtseva N, Ermolaev A, Karlov G, Kirov I, Shigyo M, Sato S, Khrustaleva L (2021) A Dual-color Tyr-FISH method for visualizing genes/markers on plant chromosomes to create integrated genetic and cytogenetic maps. Int J Mol Sci 22:5860

Kuhl J, Cheung F, Yuan Q, Martin W, Zewdie Y, McCallum J, Catanach A et al (2004) A unique set of 11,008 onion expressed sequence tags reveals expressed sequence and genomic differences between the monocot orders Asparagales and Poales. Plant Cell 16:114

Labani RM, Elkington TT (1987) Nuclear DNA variation in the genus *Allium* L (Liliaceae). Heredity 59:119–128

Lee JH, Natarajan S, Biswas MK, Shirasawa K, Isobe S, Kim HT, Nou IS et al (2018) SNP discovery of Korean short day onion inbred lines using double digest restriction site-associated DNA sequencing. PLoS One 13(8):e0201229

Li N, Qiu Z, Lu X, Shi B, Sun X, Tang X, Qiao X (2018) Comparative transcriptome analysis of temperature-induced green discoloration in garlic. Int J Genom 2018:6725728

Lutz KA, Wang W, Zdepski A, Michael TP (2011) Isolation and analysis of high quality nuclear DNA with reduced organellar DNA for plant genome sequencing and resequencing. BMC Biotechnol 11:54

Maloy S, Hughes K (2013) Brenner's encyclopedia of genetics, 2nd edn. Elsevier, New York, pp 52–54

Mandal S, Cramer CS (2021) Comparing visual and image analysis techniques to quantify fusarium basal rot severity in mature onion bulbs. Horticulturae 7:156

Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M et al (2014) Ancient hybridizations among the ancestral genomes of bread wheat. Science 345:1250092–1250092

Martin W, McCallum J, Shigyo M, Jakse J, Kuhl J, Yamane N, Pither-Joyce M et al (2005) Genetic mapping of expressed sequences in onion and in silico comparisons with rice show scant colinearity. Mol Genet Genom 274:197–204

Martínez LE, Agüero CB, López ME, Galmarini CR (2000) Improvement of in vitro gynogenesis induction in onion (Allium cepa L.) using polyamines. Plant Sci 156(2):221–226. https://doi.org/10.1016/s0168-9452(00)00263-6

Masuzaki S, Shigyo M, Yamauchi N (2006) Complete assignment of structural genes involved in flavonoid biosynthesis influencing bulb color to individual chromosomes of the shallot (*Allium cepa* L). Genes Genet Syst 81:255–263

Masuzaki S, Yahuchi S, Yamauchi N, Shigyo M (2007) Morphological characterisation of multiple alien addition lines of Allium reveals the chromosomal locations of gene(s) related to bulb formation in Allium cepa L J. Hort. Sci. Biotechnol 82:393–396

Matsuse K, Abdelrahman M, Ariyanti NA, Tsuji F, Hirata S, Nakajima T, Sato M, Hirai MY, Manochai B, Shigyo M (2022) Targeted metabolome profiling of Indonesian shallots and Japanese long-day/short-day bulb onions. Metabolites 12:1260

McCallum J, Clarke A, Pither-Joyce M, Shaw M, Butler R, Brash D, Scheffer J, Sims I, van Heusden S, Shigyo M (2006) Genetic mapping of a major gene affecting onion bulb fructan content. Theor Appl Genet 112:958–967

McCallum J, Baldwin S, Shigyo M, Deng Y, van Heusden S, Pither-Joyce M, Kenel F (2012) Allium Map-a comparative genomics resource for cultivated *Allium* vegetables. BMC Genomics 13:168

Michalik B, Adamus A, Nowak E (2000) Gynogenesis in Polish Onion Cultivars, Journal of Plant Physiology 156(2):211–216. https://doi.org/10.1016/S0176-1617(00)80308-9

Mueller RL (2015) Genome biology and the evolution of cell-size diversity. Cold Spring Harb Perspect Biol 7:a019125

Nagl W, Fusenig H-P (1979) Types of chromatin organization in plant nuclei. Plant Syst Evol Suppl 2:221–233

Ohri D, Fritsch RM, Hanelt P (1998) Evolution of genome size inAllium (Alliaceae). Pl Syst Evol 210:57–86. https://doi.org/10.1007/BF00984728

Peffley EB, Corgan J, Horak K, Tanksley S (1985) Electrophoretic analysis of Allium alien addition lines. Theor Appl Genet 71:176–184

Peška V, Mandáková T, Ihradská V, Fajkus J (2019) Comparative dissection of three giant genomes: *Allium cepa, Allium sativum*, and *Allium ursinum*. Int J Mol Sci 20:733

Rabinowicz PD, Schutz K, Dedhia N, Yordan C, Parnell LD, Stein L, McCombie WR, Martienssen RA (1999) Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. Nat Genet 23:305–308

Ranjekar PK, Pallotta D, Lafontaine JG (1978) Analysis of plant genomes V Comparative study of molecular properties of DNAs of seven *Allium* species. Biochem Genet 16:957–970

Ricroch A, Yockteng R, Brown SC, Nadot S (2005) Evolution of genome size across some cultivated *Allium* species. Genome 48:511–520

Romanov DV, Kiseleva AV, Khrustaleva LI (2014) Physical mapping of genes on the chromosomes of onion (*Allium cepa* L) using EST-clones and Tyramide-FISH. Izvestiya Timiryazev Agric Acad 1:105–114

Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M et al (2011) An integrated semiconductor device enabling non-optical genome sequencing. Nature 475:348–352

Scheffler IE (2008) Mitochondria, 2nd edn. Wiley-Liss, New York

Scholten OE, van Kaauwen MPW, Shahin A, Hendrickx PM, Keizer LCP et al (2016) SNP-markers in *Allium* species to facilitate introgression breeding in onion. BMC Plant Biol 16:187
Sekine D, Yabe S (2020) Simulation-based optimization of genomic selection scheme for accelerating genetic gain while preventing inbreeding depression in onion breeding. Breed Sci 70:594–604
Shibata F, Hizume M (2002) The identification and analysis of the sequences that allow the detection of *Allium cepa* chromosomes by GISH in the allodiploid A wakegi. Chromosome 111:184–191
Shigyo M, Kik C (2008) Handbook of plant breeding, vol 2. Springer, New York 121–159
Shigyo M, Tashiro Y, Isshiki S, Miyazaki S (1996) Establishment of a series of alien monosomic addition lines of Japanese bunching onion (*Allium fistulosum* L) with extra chromosomes from shallot (*A cepa* L Aggregatum group). Gene Genet Sys 71:363–371
Shigyo M, Iino M, Isshiki S, Tashiro Y (1997a) Morphological characteristics of a series of alien monosomic addition lines of Japanese bunching onion (Allium fistulosum L.) with extra chromosomes from shallot (A. cepa L. Aggregatum group). Genes Genet Syst 72:181–186
Shigyo M, Tashiro Y, Iino M, Terahara N, Ishimaru K, Isshiki S (1997b) Chromosomal locations of genes related to flavonoid and anthocyanin production in leaf sheath of shallot (Allium cepa L. Aggregatum group). Genes Genet Syst 72:149–152
Shigyo M, Wako T, Kojima A, Yamauchi N, Tashiro Y (2003) Transmission of alien chromosomes from selfed progenies of a complete set of Allium monosomic additions: The development of a reliable method for the maintenance of a monosomic addition set Genome 46:1098–1103
Shigyo M, Khar A, Abdelrhaman M (2018) The *Allium* genomes. In: Kole C (ed) Compendium of plant genomes. Springer, Cham
Sohn SH, Ahn YK, Lee TH, Lee JE, Jeong MH, Seo CH, Chandra R, Kwon YS et al (2016) Construction of a draft reference transcripts of onion (*Allium cepa*) using long-read sequencing. Plant Biotechnol Rep 10:383390
Song Y, Chen Y, Lv J, Xu J, Zhu S, Li M, Chen N (2017) Development of chloroplast genomic resources for *Oryza* species discrimination. Front Plant Sci 8:1854
Stack SM, Comings DE (1979) The chromosomes and DNA of *Allium cepa*. Chromosoma 70:161–181
Sulistyaningsih E, Aoyagi Y, Tashiro Y (2006) Flower bud culture of shallot (*Allium cepa* L Aggregatum group) with cytogenetic analysis of resulting gynogenic plants and somaclones. Plant Cell Tiss Org Cult 86:249–255
Sun X, Zhu S, Li N, Cheng Y, Zhao J, Qiao X et al (2020) A chromosome-level genome assembly of garlic (*Allium sativum*) provides Insights intogenome evolution and allicin biosynthesis. Mol Plant 13:1–12
Sung W, Ackerman MS, Dillon MM, Platt TG, Fuqua C, Cooper VS, Lynch M (2016) Evolution of the insertion-deletion mutation rate across the tree of life. G3 (Bethesda) 8:2583–2591
Susek A, Javornik B, Bohanec B (2002) Factors affecting direct organogenesis from flower explants of *Allium giganteum*. Plant Cell Tiss Organ Cult 68:27–33
Suzuki G, Ura A, Saito N, Do G, So B, Yamamoto M, Mukai Y (2001) BAC FISH analysis in *Allium cepa*. Genes Genet Syst 76:251–255
Tang H, Zhang X, Miao C, Zhang J, Ming R et al (2015) ALLMAPS: robust scaffold ordering based on multiple maps. Genome Biol 16:3
Tashiro Y, Tsutsumi M, Shigyo M (2000) Production and gene analyses of alien monosomic addition lines of *Allium fistulosum* L with extra chromosomes from wild species in section *Cepa* of *Allium*. Acta Hortic 521:211–218
Tsujimura M, Kaneko T, Sakamoto T, Kimura S, Shigyo M, Yamagishi H, Terachi T (2019) Multichromosomal structure of the onion mitochondrial genome and a transcript analysis. Mitochondrion 46:179–186
Tsukazaki H, Yamashita K, Yaguchi S, Masuzaki S, Fukuoka H, Yonemaru J, Kanamori H et al (2008) Construction of SSR-based chromosome map in bunching onion (*Allium fistulosum*). Theor Appl Genet 117:1213–1223
Tsukazaki H, Yamashita K, Yaguchi S, Yamashita K, Hagihara T, Shigyo M et al (2010) Direct determination of the chromosomal location of bunching onion and bulb onion markers using bunching onion-shallot monosomic additions and allotriploid-bunching onion single alien deletions. Theor Appl Genet 122:501–510

Valliyodan B, Ye H, Song L, Murphy M, Shannon JG, Nguyen HT (2017) Genetic diversity and genomic strategies for improving drought and waterlogging tolerance in soybeans. J Exp Bot 68: 1835–1849

van der Valk P, de Vries SE, Everink JT, Verstappen F, de Vries JN (1991) Pre- and post-fertilization barriers to backcrossing theinterspecific hybrid between *Allium fistulosum* L and *A cepa* L with *A cepa*. Euphytica 53:201–209

van Heusden AV, Shigyo M, Tashiro Y, van Ginkel RV, Kik C (2000) AFLP linkage group assignment to the chromosomes of *Allium cepa* L via monosomic addition lines. Theor Appl Genet 100:480–486

Vitte C, Estep MC, Leebens-Mack J, Bennetzen JL (2013) Young, intact and nested retrotransposons are abundant in the onion and asparagus genomes. Ann Bot 112:881–889

von Kohn C, Kiełkowska A, Havey MJ (2013) Sequencing and annotation of the chloroplast DNAs and identification of polymorphisms distinguishing normal male-fertile and male-sterile cytoplasms of onion. Genome 56:737–742

Vosa CG (1976) Heterochromatic handing patterns in *Allium*, II Ueterochromatic variation in species of the *Paniculatum* group. Chromosoma (Berl) 57:119–133

Vu Q H, El-Sayed M, Ito S-I, Yamauchi N, Shigyo M (2012) Discovery of a new source of resistance to Fusarium oxysporum, cause of Fusarium wilt in Allium fistulosum, located on chromosome 2 of Allium cepa Aggregatum group. Genome 55(11):797–807

Warmke HE, Lee SL (1978) Pollen abortion in T cytoplasmic male-sterile corn (*Zea mays*): a suggested mechanism. Science 200:561–563

Yaguchi S, McCallum J, Shaw M, Pither-Joyce M, Onodera S, Shiomi N et al (2008) Biochemical and genetic analysis of carbohydrate accumulation in *Allium cepa* L. Plant Cell Physiol 49:730–710

Yaguchi S, Matsumoto M, Date R, Harada K et al (2013) Biochemical analyses of the antioxidative activity and chemical ingredients in eight different *Allium* alien monosomic addition line. Biosci Biotechnol Biochem 77:2486–2488

Yamashita K (1999a) Development of cytoplasmic male sterile lines of japanese bunching onion (Allium fistulosum L.) and shallot (A. cepa L. Aggregatum group) with cytoplasms of wild species Kagoshima Univ Kagoshima, Japan PhD Thesis

Yamashita Ki, Tashiro Y (2004) Seed productivity test of CMS lines of Japanese bunching onion (Allium fistulosum L.) possessing the cytoplasm of a wild species, A. galanthum Kar. et Kir.. Euphytica 136:327–331. https://doi.org/10.1023/B:EUPH.0000032746.60805.fb

Yamashita K, Arita H, Tashiro Y (1999) Cytoplasm of a wild species, Allium galanthum Kar. et Kir., is useful for developing the male sterile line of A. fistulosum L. J Japan Soc Hort Sci 68:788–797

Yamashita K, Arita H, Tashiro Y (1999b) Isozyme and RAPD markers linked to fertility restoring gene for cytoplasmic male sterile Allium fistulosum L. with cytoplasm of A. galanthum Kar. et Kir. J Japan Soc Hort Sci 68:954–959

Yamashita K, Takatori Y, Tashiro Y (2002) Development of sequence characterized amplified region (SCAR) markers linked to the fertility restoring gene for cytoplasmic male sterile Allium fistulosum L. possessing the cytoplasm of A. galanthum Kar. et Kir. J Japan Soc Hort Sci 71:777–779

Yamashita K-I, Shigyo M, Masuzaki S-I, Yaguchi S, Hang TTM, Yamauchi N et al (2007) The application of alien-chromosome addition lines and cytoplasmic substitution lines to studies on genetics and breeding in *Allium cepa*. Genes Genomes Genom 1:27–34

Yarali F, Yanmaz R (2016) The effects of media composition on *in vitro* gynogenic embryo induction in *Allium tuncelianum* (Kollaman) Ozhatay, Matthew, Siraneci. Int J Agric Sci 8: 3247–3251

Yu N, Kim S (2020) Identification of *Ms2,* a novel locus controlling male-fertility restoration of cytoplasmic male-sterility in onion (*Allium cepa* L), and development of tightly linked molecular markers. Euphytica 217:191

Zhang C, Li X, Zhan Z, Cao L, Zeng A, Chang G, Liang Y (2018) Transcriptome sequencing and metabolism analysis reveals the role of cyanidin metabolism in dark-red onion (*Allium cepa* L) bulbs. Sci Rep 8:14109

Zhang W, Abdelrahman M, Jiu S, Guan L, Han J, Zheng T, Jia H, Song C, Fang J, Wang C (2019) VvmiR160s/VvARFs interaction and their spatio-temporal expression/cleavage products during GA-induced grape parthenocarpy. BMC Plant Biol 19:111

Eggplant (*Solanum melongena* L.) Nutritional and Health Promoting Phytochemicals

Partha Saha, Jugpreet Singh, N. Bhanushree, S. M. Harisha, Bhoopal Singh Tomar, and Bala Rathinasabapathi

Contents

Abstract

Eggplant is one of the most important Solanaceous vegetable crops which has been cultivated in different continents of the world. The greatest diversity of landraces and cultivars is found in India, China, and several countries in the Southeast Asian region. Eggplant has various culinary and medicinal uses and a great phytochemical diversity defines these uses. Phenolic compounds in eggplant are excellent antioxidants. While numerous studies on eggplant anthocyanins showed their value as antioxidative and radical-scavenging compounds, more studies on potential medicinal properties are needed. Multiple studies on the identification and mapping of genes/QTLs for the contents of chlorogenic acid, solasonine, and solamargine indicate the possibility of breeding lines

P. Saha · N. Bhanushree · S. M. Harisha · B. S. Tomar
Division of Vegetable Science, ICAR–Indian Agricultural Research Institute, New Delhi, India
e-mail: partha.saha@icar.gov.in

J. Singh
LeafWorks Inc., Sebastopol, CA, USA

B. Rathinasabapathi (✉)
Horticultural Sciences Department, University of Florida, Gainesville, FL, USA
e-mail: brath@ufl.edu

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_53

specifically improved for these phytochemicals. Metabolomics, where high-throughput identification and quantification of metabolites are done using mass spectral methods, are increasingly applied to study crop plants. Eggplant metabolites have been profiled and attempts have been made to link metabolite profiles to fruit morphology and nutrition, drought stress, nutrient use efficiency, response to pathogens, and other desirable traits. This crop has emerged as a model system for improving health benefiting traits with additional yield components. Different nuclear genomes of the eggplant have been sequenced within the last decade. With the availability of high-quality chromosome scale genomes, it is now feasible to characterize the genomic diversity and evolutionary footprints of eggplant. The information of genome sequencing is very useful to assist breeding programs to develop new varieties via marker-based selection. In this chapter we have discussed the importance of eggplants, bioactive compounds present and their role, biosynthetic pathways and progress in genome sequencing and metabolomics. We have suggested effective strategies to improve the quality of eggplant cultivars of the future.

Keywords

Anthocyanins · Chlorogenic acid · Crop breeding · Fruit quality · Nasunin · Solanum alkaloids · *Solanum melongena*

1 Introduction

Eggplant (*Solanum melongena* L.) is an important vegetable crop widely grown in many parts of the world for its immature fruit. More than 90% of production is in Asia and top producing countries are China, India, Egypt, Turkey, and Iran (Solberg et al. 2022). Eggplant and its wild relatives are grouped in the Leptostemonum clade of the Solanaceae (Weese and Bohs 2010; Särkinen et al. 2013) along with "spiny" Solanums. This crop is known as eggplant (US, Canada, Australia), brinjal (Indian subcontinent, Singapore, Malaysia, South Africa), aubergine (UK, Ireland), and Guinea squash (Southern US). The fruit of the eggplant contains numerous edible soft seeds that are bitter because of their alkaloid content. It is a rich source of vitamins, minerals, fibers, and other phytochemicals (Chlorogenic acid) which have medicinal properties (Saha et al. 2016).

Solanum melongena is native of "Indo-Chinese Center" (Taher et al. 2017) and is cultivated worldwide primarily in the tropical and subtropical regions and in temperate zone especially in the summer months or in greenhouses as off-season crop. Eggplant is known to have originated from the progenitor species *Solanum incanum* and *Solanum insanum* which have a wide distribution in India, Africa, the Middle East, and Southeast Asia (Aubriot et al. 2018; Ranil et al. 2017). A recent compilation indicated more than 19,000 accessions in various seed banks (Solberg et al. 2022). Figure 1 illustrates the species dispersal from Africa to Asia during eggplant domestication (Weese and Bohs 2010). Some of the variations in fruit shape, size, and color are shown in Fig. 2.

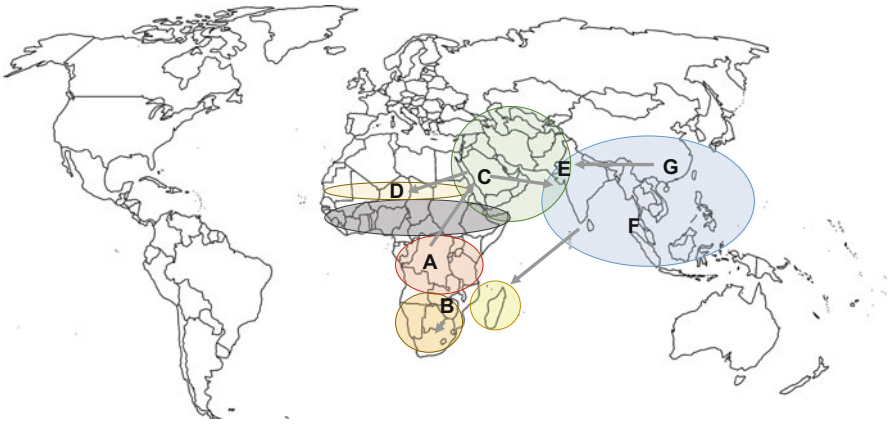


Fig. 1 Biogeography of *Solanum melongena* and related species. (Diagram adapted from Weese and Bohs (2010) and Aubriot et al. (2018)). *S. incanum* types are naturally present in regions marked A–D in eastern Africa and the Middle East and *S. melongena* types are distributed in regions marked E–H (India, China, Thailand, Korea, Vietnam, and Japan). Arrows indicate suggested directions of migration according to Weese and Bohs (2010) and Aubriot et al. (2018)

Fig. 2 Some of the variations in relative fruit size, shape, and color of immature fruit of eggplant. (Photo by B. Rathinasabapathi)

The greatest diversity of landraces and cultivars is found mainly in India, China, and countries in the Southeast Asian region. The secondary centers for this crop are in the Middle East and the Mediterranean region (Frary et al. 2007). Eggplant has been thought to be domesticated several times in Asia from *Solanum incanum*, in which human selection played a significant role. The primary traits involved in domestication process are selection for increased fruit size, decrease in prickliness, and fruit bitterness as compared to the wild progenitor species (Daunay and Hazra 2012; Meyer et al. 2012). The wild relatives *S. incanum* and another species recognized as *S. insanum* or *S. melongena insanum* have the phenotype of large spiny leaves and small, hard-textured green fruit. Species of *Solanum* that are used as

major or minor vegetables are shown in Table 1. Species of *Solanum* that are useful as sources of disease and pest resistance in eggplant breeding are shown in Table 2. A taxonomic key to the species of the eggplant clade is available in Knapp et al. (2013).

Table 1 *Solanum* species used as vegetables, their common names, distribution, and edible parts

Species	Common name	Distribution	Edible part(s)
S. aethiopicum	Molk tomato Golden apple	Tropical Africa	Leaves and immature fruit
S. americanum	Glossy nightshade	Central Eastern USA	Leaves and green fruit
S. gilo	Gilo, Jilo	Western Africa, Brazil	Fruit, shoot
S. hirsutissimum	Lulita	Central and South America	Mature fruit
S. incanum	Sodom apple	West Africa, India	Immature fruit
S. indicum	Indian nightshade	Southeast Asia	Fruit, leaves
S. integrifolium	Scarlet eggplant	Warm temperate, Subtropical Asia	Fruit
S. lycopersicum	Tomato	American Tropics	Fruit
S. macrocarpon	African eggplant	Subtropical and tropical regions, West and Central Africa	Fruit, leaves
S. melanocerasum	Garden Huckleberry, Wonderberry, Black nightshade	Warm temperate regions, Subtropical America and West Africa	Mature fruit
S. melongena	Eggplant, brinjal, aubergine	Warm temperate, Subtropical and tropical regions	Immature fruit
S. muricatum	Pepino, melon pear	Subtropical and tropical America	Mature fruit
S. quitoense	Naranjillo, lulo	Subtropical and tropical America	Mature fruit
S. sessiliflorum	Cocona	Northern South America	Mature fruit
S. torvum	Turkey berry	South American tropics	Immature fruit
S. tuberosum	Potato	Central and South America	Tuber

Table 2 Wild *Solanum* species as sources of resistance to pests and diseases (Swarup 1995)

Species	Source of resistance
S. melongena var. *insanum, S. integrifolium*	Bacterial wilt
S. incanum, S. indicum, S. integrifolium	Fusarium wilt
S. indicum, S. aculeatissimum	Verticillium wilt
S. xanthocarpum, S. nigrum, S. sisymbrifolium	Fruit rot
S. torvum, S. mammosum, S. khasianum	Spotted beetle
S. sisymbrifolium, S. aethiopicum	Root knot nematode
S. viarum, S. incanum, S. gilo	Little leaf

2 Eggplant as a Source of Food, Nutrition, and Health Promoting Compounds

Eggplant has various culinary and medicinal uses. Here, we summarize a variety of culinary uses for immature fruit of eggplant from different areas of the world. Eggplant is mainly used to prepare *Baingan ka Bhartha* or *Gojju*, stir fries and soups in India, *Yu Xiang Qiezi* and *Di San Xian* stir fries in the Chinese cuisine, *nasu dengaku* of Japan, *ratatouille* of the French cuisine, *moussaka* and *melitzanosalata* of the Greek, *baba ghanoush* of the Middle East, *nigvzianibadrijani* in the Caucasus and the *melanzaneallaparmigiana* of the Italian cuisine. Two species namely, *S. aethiopicum* and *S. macrocarpon* are commonly grown for the consumption of cooked leaves in Africa. Eggplant is a "low-calorie vegetable" due to its low energy density. But the presence of vitamins, minerals, proteins, fiber, and phenolic compounds in the fruit proves its beneficial value to human health (Frary et al. 2007). The nutritive value of eggplant is presented in Table 3. A comparative study on tomato, pepper, and eggplant showed that eggplant had greater levels of mineral nutrients potassium, magnesium, and copper, total sugars, and the antioxidant chlorogenic acid compared to pepper and tomato (Rosa-Martínez et al. 2021).

Besides, as a source of food, eggplant is used in traditional medicine to treat asthma, bronchitis, cholera and dysuria, skin-related problems, and in lowering blood cholesterol levels (Rotino et al. 2014; Meyer et al. 2014). Studies in rats suggested that eggplant could reduce the absorption of dietary cholesterol (Kritchevsky et al. 1975). However, small scale clinical trials on the potential of dietary eggplants in reducing serum lipid levels have shown either no or transitory effects only (Guimaraes et al. 2000; Praca et al. 2004). Donmez et al. (2020) showed hemorrhoid healing activity for the extracts of eggplant calyx in an animal model. Additional controlled clinical studies are needed to test efficacy and modes of action.

As other Solanaceous crops tomato, potato, and pepper, eggplant also contains glycoalkaloids such as solasonine and solamargine, which are known to have beneficial biological and anticancer properties (Cham 2012). However, these alkaloids at excess levels exert potential toxic effects (Chami et al. 2003) and impart bitterness to fruits making them unpalatable (Frary et al. 2007). But the cultivated varieties of *S. melongena* have levels of alkaloids which are acceptable for human consumption, unless the plants have been subjected to stress conditions. Figure 3 depicts structures of major health promoting phytochemicals in eggplant.

Several studies have documented that the *Solanum* alkaloid solamargine induces apoptosis of cancer cells. Via multiple signaling pathways, solamargine sensitizes the cells to certain cancer drugs, and hence may have promising applications for cancer therapy (Kalalinia and Karimi-Sani 2017). Solamargine has also been reported to have anti-inflammatory effects against UVB-induced skin hyperpigmentation (Zhao et al. 2022).

Phenolic compounds in eggplant are excellent antioxidants and there is high genetic variation for this trait (Nandi et al. 2021; Plazas et al. 2013a, b). In addition to their antioxidant potential eggplant extracts were shown to inhibit alpha-glucosidase and angiotensin converting enzyme, target enzymes for therapies for

Table 3 Nutritional profile of raw eggplant (per 100 g fresh weight). Values from the USDA "Food Data Central" (Haytowitz et al. 2019) are compared to mean and standard error values from a study that examined ten eggplant varieties grown under same horticultural conditions (Rosa-Martínez et al. 2021)

Composition	Raw	Mean and standard error for ten varieties
Proximate		
Water (g)	92.3	90.07 + 27.2
Energy (kcal)	25	NA
Protein (g)	0.98	1.49 ± 0.37
Total lipid (g)	0.18	NA
Carbohydrate, by difference (g)	5.88	NA
Total sugars (g)	3.53	3.19 ± 0.83
Total dietary fiber (g)	3	NA
Minerals		
Calcium (mg)	9	10.29 ± 2.3
Iron (mg)	0.23	0.35 ± 0.11
Magnesium (mg)	14	17.93 ± 4.66
Phosphorus (mg)	24	38.89 ± 7.79
Potassium (mg)	229	326.64 ± 56.71
Sodium (mg)	2	4.10 ± 1.52
Zinc (mg)	0.16	0.30 ± 0.08
Vitamins		
Vitamin C (mg)	2.2	5 ± 1.0
Niacin (mg)	0.649	NA
Pantothenic acid (mg)	0.281	NA
Folate, total (μg)	22	NA
β-Carotene (μg)	14	48 ± 52
Vitamin A (IU)	23	NA
Vitamin E (α-tocopherol) (mg)	0.3	NA
Vitamin K (phylloquinone) (μg)	3.5	NA
Lipids		
Fatty acids, total saturated (g)	0.034	NA
Fatty acids, total monounsaturated (g)	0.016	NA
Fatty acids, total polyunsaturated (g)	0.076	NA
Cholesterol (mg)	0.000	NA

type 2 diabetes and hypertension respectively (Kwon et al. 2008). Among the wild accessions, *Solanum insanum* has significantly higher phenolic content than *Solanum xanthocarpum* and *Solanum khasianum*. The wild accessions are rich sources of total phenolics (Kaur et al. 2014; Kaushik et al. 2020; Nandi et al. 2021).

Early research in eggplant breeding focused on the improvement of horticultural characteristics such as fruit yield, fruit size, color, and shape (Kashyap et al. 2003; Frary et al. 2007), low prickliness, and adaptation to climatic conditions (Daunay et al. 2001). A few studies focused on the resistance to biotic stresses (Rotino et al. 2014). Within the last few years, however, there is an increased research interest on

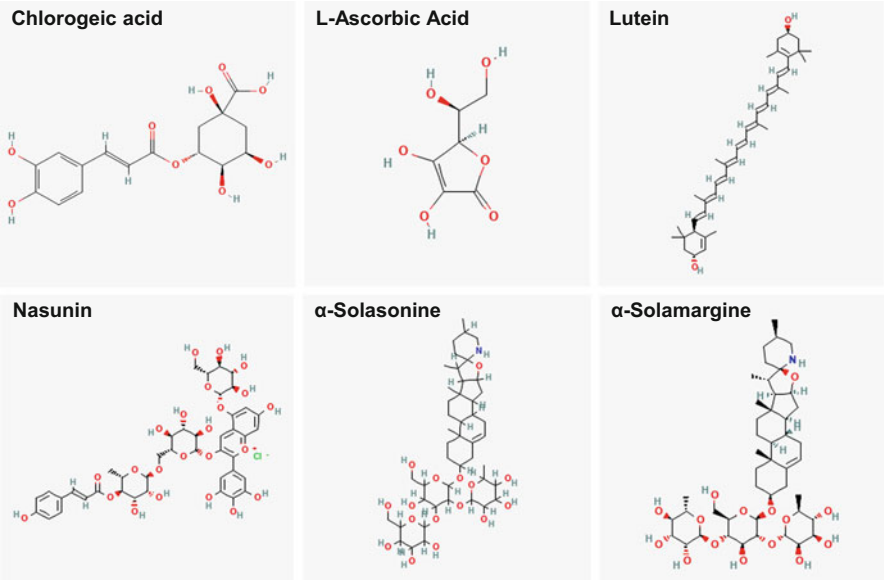


Fig. 3 Health promoting phytochemicals in eggplant. Representative compounds from the organic acids, phenolic compounds, carotenoids, anthocyanins and glycoalkaloids are shown. Note that other forms of these compounds and related compounds with bioactivity are also known but not depicted. Structures are from PubChem database

organoleptic and nutritional properties, bioactive metabolites, and postharvest traits of eggplant fruit. A few of these studies also highlight genetic and genomic perspectives (Prohens et al. 2007, 2013; Gajewski et al. 2009; Plazas et al. 2013a, b; Zhang et al. 2016; Docimo et al. 2016b; Scalzo et al. 2016; Mangino et al. 2022).

Eggplant is a nutritionally valuable vegetable as a good a source of vitamins and minerals (Grubben et al. 1977, Table 3). It also contains several classes of health-promoting metabolites including anthocyanins (delphinidin glycosides) and chlorogenic acid (Stommel and Whitaker 2003; Mennella et al. 2010) with nutraceutical and antioxidant properties (Cao et al. 1996; Kwon et al. 2008; Akanitapichat et al. 2010). An eggplant variety 'Pusa Hara Baingan 1' developed from ICAR-IARI, New Delhi has fruit with high antioxidant activity (3.41 CUPRAC μ mol trolox /g, 3.07 FRAP μ mol trolox /g) (Kumar et al. 2020). The bioactive compounds and health promoting properties of eggplant is presented in Table 4. Eggplant also contains certain anti-nutritional compounds, like toxic saponins and steroidal glycoalkaloids (SGA), which form the basis for the bitter taste of the flesh (Aubert et al. 1989a, b; Sánchez-Mata et al. 2010). The accumulation of these metabolites in the fruit and the total amount of all these metabolites appear to be influenced by genetic and developmental factors (Mennella et al. 2012). There is genetic variability for phytochemical accumulation traits among different eggplant varieties and environmental stress affects these traits (Plazas et al. 2013a, b).

Multiple studies focused on the genetic basis for fruit quality traits in eggplant. Quantitative trait loci (QTLs) affecting the content of health promoting or

Table 4 Reports of bioactive compounds and health promoting properties of eggplant

Compound	Potential health benefits	References
Chlorogenic acid	Antioxidant, anti-inflammatory, cardioprotective, antidiabetic, antimicrobial, and neuroprotective	Plazas et al. (2013a)
Nasunin	Antioxidant activity, neuroprotective, cardiovascular protection, anticancer, anti-diabetic, anti-inflammatory	Casati et al. (2016), Matsubara et al. (2005), Lin and Li (2017), and Wang and Stoner (2008)
Solasodine	Anticancer, anti-inflammatory	Shen et al. (2017), da Costa et al. (2015), and Friedman (2006)
Fiber content	Digestion, anticancer against colon cancer	Fraikue (2016)
Polysaccharides	Immuno modulation, antitumor effects, and antioxidant	Mei et al. (2017)
Delphinidin	Reduction of oxidative stress and vascular inflammation	Watson and Schönlau (2015) and Harisha et al. (2023)
Kaempferol	Antioxidant defense and a reduction of the risk of chronic diseases, especially cancer	Chen and Chen (2013)
Myricetin	Antioxidant defense, anti-carcinogenic, antimicrobial, and antiplatelet	Li and Ding (2012)
Quercetin	Antioxidant defense, cytoprotective effects, antimicrobial, anti-inflammatory, and muscle-relaxing properties	Jan et al. (2010)
Luteolin	Antioxidant defense, anti-inflammatory, properties	Jiang et al. (2013)
Isorhamnetin	Antioxidant and antitumor activity on human cancer cells, prevention of endothelial cell injuries caused by oxidized low-density lipoprotein	Jaramillo et al. (2010)
Lutein	Non-nutritive carotenoid, antioxidant in the retina, protecting the eye from inflammatory damage by light	van Lent et al. (2016)
Zeaxanthin	Anti-inflammatory effects via reducing oxidative damage of the retina	Manikandan et al. (2016)
β-Cryptoxanthin	Vitamin A precursor, may help reduce free radical damage to biomolecules, anticancer	Lorenzo et al. (2009)
Tannins	Enhance glucose uptake and inhibit adipogenesis. Inhibition of LDL-cholesterol oxidation	Kumari and Jain (2012)
Hydroxycinnamic acids	Free radical-scavenging properties, protection from side effects of chemotherapy	El-Seedi et al. (2012)

antinutritional compounds are known. These studies have also identified QTLs for various other fruit quality traits (Shetty et al. 2011; Gramazio et al. 2014). QTLs associated to fruit content of sugars, organic acids, dry matter and total soluble solids, chlorogenic acid, the steroidal glycoalkaloid solamargine, pigments delphinidin-3-

rutinoside (D3R), and delphinidin-3-(p-coumaroylrutinoside)-5-glucoside (nasunin) are useful resources to breed cultivars improved for specific phytochemical profiles. Very recently, the HPLC analysis identified cyanidin-3-*O*-glucoside, delphinidin-3-*O*-glucoside and delphinidin-3-rutinoside anthocyanins in brinjal lines and the highest delphinidin-3-*O*-glucoside was found in purple fruited variety Pusa Upkar, delphinidin 3-rutinoside in purple fruited line BR-40-7 (Harisha et al. 2023).

3 Major Bioactive Nutraceutical Compounds Present in Eggplant

3.1 Polyphenolics

Among all the studied Solanaceous crops, eggplant is a rich source of total phenolic acids, which are nutritionally important bioactive constituents (Helmja et al. 2007). The most prevalent group of phenolic compounds present in eggplant are the conjugates of hydroxycinnamic acid (HCA) and their concentration ranges from 4240 to 9610 mg/kg. HCA conjugates are synthesized by converting phenylalanine to cinnamic acid by the enzyme phenylalanine ammonia-lyase (PAL). Among HCA conjugates, chlorogenic acid (5-*O*-caffeoyl-quinic acid; CGA) an ester of HCA, contributes to 70% to over 95% of the total phenolic content (Whitaker and Stommel 2003; Plazas et al. 2013a; Niño-Medina et al. 2017). Quinic acid has four stereocenters namely carbons 1, 3, 4, and 5, and each one of these positions can be substituted by caffeic acid giving rise to 15 possible combinations. While CGA is the major hydroxycinnamic acid conjugate in eggplant, its 3-*O*, 4-*O*, 5-*O*-cis isomers, and 3,5- and 4,5-dicaffeoylquinic acid isomers have also been identified (Whitaker and Stommel 2003). Chemistry, synthesis, analytical challenges and bioactivity of caffeoylquinic acids in plants have recently been reviewed by Magania et al. (2021).

Growing interest in breeding for high CGA is mainly due to its beneficial properties including antioxidant (Chen et al. 2009; Zhao et al. 2012), anti-inflammatory (Morishita and Ohnishi 2001; dos Santos et al. 2006; Jin et al. 2006; Sheu et al. 2009; Sato et al. 2011), anticarcinogenic, antimicrobial (Almeida et al. 2006), anti-obesity (Cho et al. 2010), antipyretic, neuro-protective and analgesic properties (Cho et al. 2010; Plazas et al. 2013a; dos Santos et al. 2006). As a source, eggplant contributes substantially to CGA intake and there exists a significant variability among eggplant lines studied as compared to other sources of fruits and vegetables (Chumyam et al. 2013; Kaur et al. 2014; Prohens et al. 2013; Okmen et al. 2009; Chioti et al. 2022). The biochemical pathway and range of CGA content in different eggplant lines in comparison to other sources are shown in Fig. 4 and Table 5, respectively.

Fruit CGA content of eggplant fruit is influenced by both genetics and environmental factors including the fruit developmental stage, season, and storage conditions (Plazas et al. 2013a). The crop grown in the summer season showed a trend of decrease phenolic compounds compared to spring-grown crop suggesting negative influence of high temperatures on phenolic content. This information will help in monitoring the suitable time of harvest (García-Salas et al. 2014). In some of the

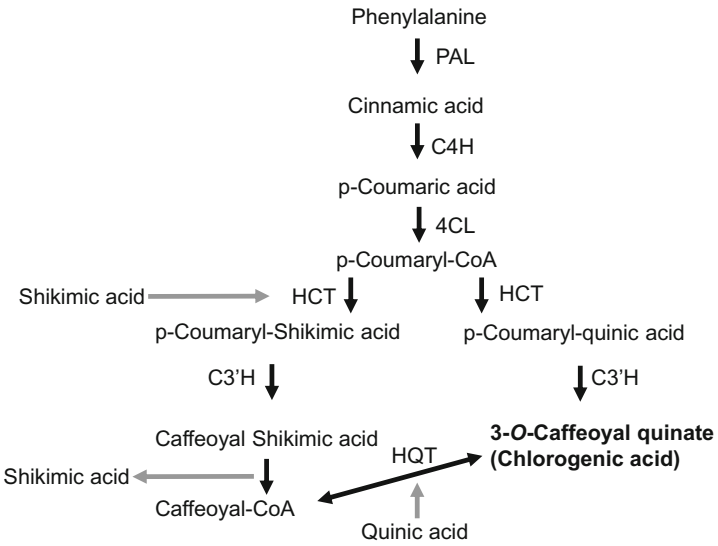


Fig. 4 Biosynthetic pathway of chlorogenic acid in eggplant. Enzyme catalyzed steps are shown with an arrow and the names of the enzymes are shown in three-letter abbreviations of enzyme names listed in Table. (Adapted from Plazas et al. (2013a))

Table 5 Chlorogenic acid contents of eggplant compared with other major vegetables (as summarized by Plazas et al. 2013a)

Plant source	CGA (g kg^{-1} dw)	References
Eggplant	4.9–21.6	Stommel and Whitaker (2003)
	4.2–9.5	Whitaker and Stommel (2003)
	1.5–2.2	Gajewski et al. (2009)
	5.0–8.1	Singh et al. (2009)
	2.6–6.7	Luthria et al. (2010)
	11.2–24.0	Mennella et al. (2010)
	1.4–8.4	Luthria et al. (2010)
	14.1–28.0	Mennella et al. (2012)
Carrot	0.3–18.8	Sun et al. (2009)
Tomato	0.2–0.4	Hallmann (2012)
Pepper	0.7–0.9	Hallmann and Rembiałkowska (2012)
Artichoke	1.1–1.8	Lutz et al. (2011)

studies it was also noticed that organically grown eggplants had greater levels of total phenolics than conventionally produced eggplants (Raigon et al. 2010), but the contrary report also exists in which conventional and organic cultivation of the American variety 'Blackbell' had comparable levels of phenolic contents (Luthria et al. 2010). As compared to other derivatives of phenolic compounds CGA is quite stable at high temperatures, therefore boiling eggplant boosts its bioavailability when compared to uncooked eggplant (Lo Scalzo et al. 2010).

3.1.1 Genetic Sources of Phenolic Compounds

Generally, wild relatives had higher content of total phenolics with broader range of variation in them as compared to cultivated eggplant (Kaushik et al. 2017). The results of screening of the eggplant germplasm for the sources of high CGA content indicate that wild species *S. incanum* can contribute significantly for improving the CGA content in eggplant. Other species, like *S. sodomaeum* L. (*S. linneanum* Hepper & Jaeger) also showed a higher CGA content than that of *S. melongena* (Mennella et al. 2010). Therefore, utilization of wild relatives, such as *S. incanum*, in breeding of eggplant varieties will be of great interest. Along with the wild species the landraces also represent another useful source of variation. Introgressive breeding and selection for desired profiles of phenolic compounds along with other horticulturally relevant traits should be useful in this regard.

3.1.2 Breeding Strategies for Increased CGA

The strategies proposed for breeding of high CGA content exploit naturally available variation to selection among accessions, selection from intraspecific variation, development of hybrids and inbred lines or introgression of the high CGA trait from wild species into cultivated background. It is also necessary to have good horticultural and commercial characteristics along with the improved concentrations of CGA in new cultivars (Daunay 2008). Molecular markers for traits associated with high CGA content will be useful for developing new and improved cultivars in a resource-efficient manner. In case of conventional breeding methods, utilization of intraspecific variation and selection among the accessions or varieties resulted in the identification of lines with increased content of CGA (Stommel and Whitaker 2003; Whitaker and Stommel 2003; Hanson et al. 2006; Prohens et al. 2007; Raigón et al. 2008; Okmen et al. 2009; Mennella et al. 2012). An alternative and very successful method of breeding for high CGA is to develop hybrids along with commercial traits by using the lines having parents with high CGA content. The developed hybrids can be further used to select and develop inbred lines with higher content in CGA as suggested by the researchers (Prohens et al. 2007: Raigón et al. 2008). Prohens et al. (2013) improved the phenolic acid content in commercial eggplant lines by developing interspecific *S. melongena* and *S. incanum* backcross generation (BC1) progenies to introgress the alleles of *S. incanum* involved in CGA biosynthesis in the genetic background of *S. melongena*. *S. incanum* was found fully fertile with *S. melongena*. The individuals having high content in CGA in the first backcross generation was found and selected. The gene action studies from the population suggested the presence of additive genetic effects explaining CGA variation. Therefore, it is necessary to have the alleles of *S. incanum* in homozygous state to obtain lines with stable and high levels of CGA in the fruit.

The identification and mapping of genes involved in the CGA synthesis pathway was carried out by Gramazio et al. (2014). The authors were successful in amplifying the candidate genes encoding enzymes of CGA synthetic pathway and the identified genomic regions were syntenic with the tomato genome (Wu et al. 2009). Table 6 lists the candidate genes and their linkage group locations in the eggplant genome as reported by Gramazio et al. (2014).

Table 6 Candidate genes encoding enzymes of chlorogenic acid (CGA) biosynthesis and their chromosomal locations

S. No	Candidate genes coding for biosynthetic enzymes	Linkage group
1	*PAL* (phenylalanine ammonia-lyase)	9
2	*C4H* (cinnamic acid 4-hydroxylase),	6
3	*4CL* (4-coumaroyl: CoA-ligase)	3
4	*C3H* (p-coumarate 3-hydroxylase),	1
5	*HQT* (hydroxycinnamoyl CoA quinate hydroxycinnamoyltransferase),	7
6	*HCT* (hydroxycinnamoyl-coAshikimate/quinatehydroxycinnamoyltransferase)	3

Gene transfer technologies are well developed in eggplant and have been used to improve several traits (Acciarri et al. 2000; Donzella et al. 2000; Pal et al. 2009). Kaushik et al. (2020) obtained transgenic plants overexpressing the central enzyme hydroxycinnamoyl CoA-quinate transferase (*SmHQT*), which catalyzes the reaction to the chlorogenic acid synthesis. Using agro infiltration technique the gene construct having *SmHQT* gene was transformed into the eggplant and was further validated by HPLC analysis of transgenic tissue for CGA. The results of the study confirmed the increase in CGA content by twofold in the transformed plants. When the general public's acceptance of genetically modified crops (Raybould and Poppy 2012) improve and national policies toward genetically modified crops become favorable, nutritionally enhanced transgenic eggplant cultivars will become more widely useful.

3.1.3 Effect of CGA Content on Fruit Flesh Browning

When an eggplant fruit is cut, the phenolic compound CGA present in eggplant fruits acts as the substrate for polyphenol oxidases which catalyze the oxidation of phenols to quinones. Quinones reacting non-enzymatically with O_2 and other molecules cause tissue browning (Ramírez and Virador 2002; Prohens et al. 2007; Plazas et al. 2013b; Mishra et al. 2013; Prohens et al. 2013; Docimo et al. 2016a).

The polyphenol oxidase reaction is a major drawback for increasing the CGA content in fruit (Prohens et al. 2007, 2013; Plazas et al. 2013b; Mishra et al. 2013) because browning reduces the visual quality of the fruit both for the fresh market and for the processing industry (Mishra et al. 2013). In order to reduce this potential side effect, it has been emphasized for indirect selection for lower content of total phenolic compounds in fruit flesh. Also, variation for PPO activity was observed between the different groups of the germplasm. The primary gene pool species consisting of cultivated eggplant and hybrids display significantly lower PPO specific activities than those of the wild species of the secondary and tertiary gene pools (Kaushik et al. 2017). Eggplant's wild relatives had higher PPO activity than the cultivated species (Shetty et al. 2011), suggesting enhanced defense in the wild species. In several crops, PPO activity was shown to have a positive correlation with browning traits, which indicates that by selecting for low PPO activity, it is possible to develop materials with reduced browning (Di Guardo et al. 2013; Nayak et al. 2015;

Urbany et al. 2011). In eggplant, the correlation between fruit flesh phenolic concentration, and particularly CGA content, and fruit flesh browning has been found to be moderate (Prohens et al. 2007; Plazas et al. 2013a; Docimo et al. 2016a). Other physiological or cell anatomical factors related to flesh browning were suggested (Docimo et al. 2016a; Mishra et al. 2013; Prohens et al. 2007). Prohens et al. (2007) found a moderate relationship ($r = 0.389$) between the total phenolic content and fruit flesh browning among the cultivated eggplants. This study suggests that there is a possibility to select varieties with high content in phenolics but with low or moderate fruit flesh browning traits. In the later studies by Prohens et al. (2013), fruit flesh browning and total content of hydroxycinnamic conjugates (CGA) has been correlated in interspecific progenies between *S. melongena* and *S. incanum.* The genetic and QTL mapping studies carried out by Gramazio et al. (2014) led to identification and mapping of five genes that encode PPO enzymes in eggplant. All the five genes *PPO1, PPO2, PPO3, PPO4*, and *PPO5* were mapped on the linkage group 8 and were found to be syntenic with the tomato genome (Wu et al. 2009). Presence of all the five genes on same linkage group suggest that *PPO* genes form a cluster in eggplant genome (Newman et al. 1993; Thipyapong et al. 2007; Tran et al. 2012) and it was also confirmed that the genes involved in CGA synthetic pathway were not linked to the *PPO* gene cluster.

3.2 Anthocyanins

Anthocyanins are the most abundant class of water-soluble pigments in plants (Mazza 2007). They are mainly responsible for red, purple, and blue colors in fruits, vegetables, flowers, and grains. Derivatives of six anthocyanidins, namely pelargonidin, cyanidin, delphinidin, peonidin, petunidin, and malvidin are widespread in plants (Kong et al. 2003). Among the above mentioned anthocyanidins, delphinidin glycosides [i.e., delphinidin-3-rutinoside (D3R) and delphinidin-3-(p-coumaroylrutinoside)-5-glucoside (nasunin)] were found more abundantly in eggplant than any other groups (Toppino et al. 2016). Multiple studies have identified several delphinidin derivatives in eggplant accessions and they have been reviewed by Niño-Medina et al. (2017).

Anthocyanins are found to have significant biological functions in plants as insect pollinator attractants, seed dissemination agents and protector of plants against various biotic and abiotic stresses (Chalker-Scott 1999; Ahmed et al. 2014). They were found to have photo protective activity, by absorbing excess visible and UV light and scavenging free radicals thus protecting the photosynthetic apparatus (Guo et al. 2008). Anthocyanin extracts have been shown to have antioxidative and radical-scavenging activities (Astadi et al. 2009), and serve as chemoprotective agents. Anthocyanins also have anti-diabetic characteristics, such as decreasing cholesterol (DeFuria et al. 2009), and increasing insulin secretion (Matsui et al. 2004). There are lines of evidence that a balanced diet rich in anthocyanin-rich foods can significantly reduce the chance of developing various chronic illnesses, including cancer (He and Giusti 2010). While certain health-promoting effects of

anthocyanins could be due to their high antioxidant activities, specific anthocyanins can modulate signaling pathways in mammalian cells, which could explain some of their specific beneficial effects (Salehi et al. 2020).

Light and temperature are the main environmental factors regulating the anthocyanin biosynthetic pathway. In general, high light intensity increases anthocyanin production in many plant species (Maier and Hoecker 2015). In one study addition of UV-A irradiation improved anthocyanin pigmentation of eggplants when grown in a greenhouse with low UV transmittance (Matsumaru et al. 1971). It is well known that low temperature stress induced anthocyanin accumulation in plants of the Solanaceae (Løvdal et al. 2010; Jiang et al. 2016). It has been proposed that the promotion of anthocyanin accumulation by low temperature and light might be through the same or overlapping mechanism (Jaakola 2013; Xu et al. 2015), as induction of anthocyanin biosynthesis at low temperature needed light.

3.2.1 Biosynthetic Pathway of Anthocyanin

Malonyl-CoA and 4-Coumaroyl-CoAarethe precursors for the flavonoid biosynthetic pathway in plants. Chalcone synthase (*CHS*) catalyzes the first committed step. The anthocyanin biosynthesis process is a conserved network which is well studied in crop plants. Figure 5 illustrates the steps leading to delphinidin, the most abundant anthocyanin in the eggplant.

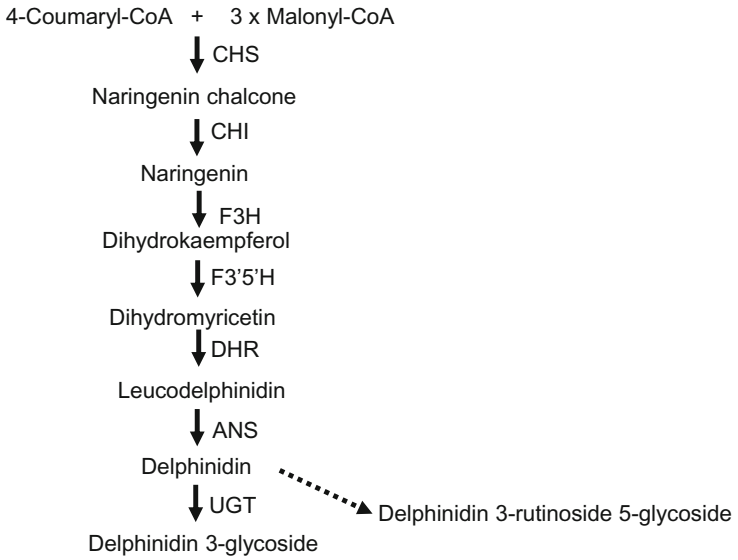


Fig. 5 Key steps in the biosynthesis of anthocyanin nasunin (delphinidin 3-rutinoside-5-glucoside) based on *KEGG* pathway database's anthocyanin biosynthesis. Steps with dotted arrows denote uncharacterized steps, but beginning to be understood (Florio et al. 2021). Enzyme name abbreviations are *CHS* chalcone synthase, *CHI* chalcone isomerase, *F3H* hydroxylase, *F3′,5′H* hydroxylase, *DHR* dihydromyricetin reductase, *ANS* anthocyanin synthase, *UGT* glycosyltransferase

The early biosynthetic genes (*EBGs*) are known to involve in genetic regulation of anthocyanin biosynthesis. The EBGs namely – *CHS*, *CHI*, and *F3H* are the common flavonoid biosynthetic pathway genes which are involved in the synthesis of all flavonoids and late biosynthetic genes (*LBGs*) namely – *F3′H, F3′ 5′H, DFR, ANS*, and *UFGT* are required for the biosynthesis of specific classes of flavonoids, including anthocyanins. In eggplant, *EBGs* (*SmCHS, SmCHI,* and *SmF3H*) were reported to respond to low temperature and light earlier than LBGs (*SmF3′ 5H, SmDFR*, and *SmANS*) (Jiang et al. 2016).

Positive correlations between expression levels of LBGs and anthocyanin content have been recorded in many Solanaceous vegetables (Borovsky et al. 2004; André et al. 2009; Povero et al. 2011; Aza-Gonzalez et al. 2013). *MYB-bHLH-WD40* (*MBW*) complex regulates the structural genes of the anthocyanin biosynthetic pathway. This complex consists of MYB, basic helix-loop-helix (*bHLH*) and *WD40* repeat family proteins. A study combining transcriptomic and proteomic analyses on light-induced anthocyanin biosynthesis in eggplant (Li et al. 2017) identified multiple genes and gene products critical in the accumulation of anthocyanins. It should be noted that although much is known about the pathway, phytochemistry of various forms of delphinidin derivatives in eggplant and the enzymatic steps leading to them are poorly known.

3.2.2 Breeding for Anthocyanin Content in Eggplant

In recent days, consumers prefer anthocyanin-rich food products, and such products are marketed for their health promoting potentials (Pojer et al. 2013). The correlation study between the antioxidant activity and total anthocyanin concentration in different eggplant types was carried out by Nisha et al. (2009). The extracts from small sized purple-color fruit showed greater antioxidant activities with maximum concentration of anthocyanin (0.756 mg/100 g) than that of other samples, such as long green (0.0475 mg/100 g), purple colored moderate size (0.525 mg/100 g) and purple colored big size (0.553 mg/100 g). Similar study was carried out by Dhruve et al. (2014) to screen the available eggplant genotypes and to quantify the amount of anthocyanin present. In those studies, AB-07-02 was found to have maximum anthocyanin content in peel of about 474.85 mg/100 g and in the genotypes GBL-1 and GP-White anthocyanin content was not detected. Screening of 26 different colored eggplant genotypes by Koley et al. (2019) reported that the total anthocyanins (in monomeric form) in the purple genotype varied from 27.63 to 359.28 μg C3G/100 g FW, whereas in green and white-colored genotypes anthocyanin was not detected.

It is important to know the inheritance pattern and nature of gene action for selection of breeding methods to improve a particular trait. A study was conducted to decipher the inheritance pattern for anthocyanin using the F_2 and backcross population from a cross of 'Pusa Safed Baingan 1' and 'Pusa Uttam' implying dominant epistasis (Bhanushree et al. 2019). Efforts have been made to develop purple fruited varieties in India. Few dark purple fruited varity Pusa Oishiki, Pusa Vaibhav, Pusa Unnat (F1) rich in anthocyanin were developed very recently (Saha et al. 2021, 2022, 2023). Barchi et al. (2012) identified QTLs determining anthocyanin

pigmentation in different plant parts of eggplant by conducting the experiment in two different locations. This study looked for marker associations to the following phenotypes: adaxial leaf lamina anthocyanin (*adlan*), stem anthocyanin (*stean*), abaxial leaf lamina anthocyanin (*ablan*), calyx anthocyanin (*calan*), corolla color (*corcol*), leaf venation anthocyanin (*lvean*), and fruit peduncle anthocyanin (*pedan*). The list of QTLs detected is shown in Table 7. This study revealed that QTLs located in five different chromosomes control anthocyanin accumulation in this population. Later, QTL studies by Toppino et al. (2016) identified the QTLs affecting fruit color traits and nasunin. Two QTLs for fruit color traits were mapped on chromosome 5, explaining 56.3% and 69.9% of the phenotypic variance in two different locations respectively. One QTL was detected on the same chromosome (chromosome number 5) explaining 28% and 28.4% of the variance in two different locations.

In a different study eggplant was genetically engineered by introducing *SmMYB1* gene encoding a *R2R3 MYB* transcription factor involved in regulating anthocyanin biosynthesis in the peel of eggplant into a non-anthocyanin-accumulating cultivar (*Solanum aethiopicum* group Gilo) via *Agrobacterium*-mediated transformation (Zhang et al. 2016). The transgenic plants had high concentrations of anthocyanin

Table 7 QTLs detected in the mapping population by Barchi et al. (2012), their chromosomal location (Chrom), log of odds ratio score (LOD) for detection of the QTL, percent contribution to phenotypic variance (PVE), and additive (A) and dominant variance (B) contributions

Trait code	QTL	Chrom	Position (cM)	LOD	PVE	A	D
Adlan	*adlanE06.M*	6	151.48	7.93	8.00	−20.282	0.287
	adlanE10.M	10	69.39	36.98	60.60	−20.948	0.060
Stean	*steanE05.ML*	5	69.73	14.59	14.80	−20.251	0.234
	steanE10.ML	10	68.92	36.60	53.60	−20.487	0.252
Ablan	*ablanE06.ML*	6	151.482	8.21	8.70	−20.209	0.259
	ablanE10a.ML	10	68.92	29.89	45.20	−20.642	−20.004
	ablanE10b.ML	10	0	6.59	6.80	−20.255	20.101
	ablanE11.ML	11	83.275	4.51	4.50	−20.177	−0.227
Calan	*calanE05.ML*	5	75.30	12.39	8.80	−20.213	0.147
	calanE06.ML	6	151.48	4.48	2.70	−20.094	0.117
	calanE08.ML	8	27.53	5.11	3.40	−20.112	0.130
	calanE10.ML	10	68.92	47.53	61.00	−20.553	0.290
Corcol	*corcolE05.ML*	5	75.30	34.08	63.70	−21.545	1.490
Lvean	*lveanE05.ML*	5	75.30	13.71	7.80	−20.274	0.189
	lveanE10.ML	10	69.13	58.94	73.90	−20.833	0.398
Pedan	*pedanE01.ML*	1	118.30	4.69	1.50	0.067	−0.191
	pedanE05.ML	5	59.81	5.95	1.90	−20.110	0.227
	pedanE10a.ML	10	69.13	73.20	76.40	−21.029	0.551
	pedanE10b.ML	10	0.00	6.33	2.00	−20.191	−20.039
	pedanE12a.ML	12	106.73	7.12	2.30	−20.190	0.020
	pedanE12b.ML	12	30.23	5.73	1.80	0.181	0.032

in leaves, petals, stamens, and fruit peels under normal growth conditions, especially in fruit flesh. Later studies of qRT-PCR analysis revealed that transformed plants showed significantly high up-regulation of anthocyanin structural genes than the non-transformed plants. Additionally, they had a greater tolerance to freezing stress and better recovery under rewarming conditions, suggesting that anthocyanin synthetic and regulatory circuits are potential targets to improve eggplants for stress tolerance. Recently, the proof for *R2R3-MYB* transcription factor *SmMYB75* promotes anthocyanin biosynthesis in eggplant was supported by the study of Shi et al. (2021). Overexpression of *SmMYB75* in the eggplant cultivar '116' showed increased accumulation of anthocyanins, which resulted in a change in the color of the calli from green to purple. The Yeast one-hybrid (*Y1H*) and dual-luciferase (*Dual-LUC*) assays showed that *SmMYB75* could bind to the promoter of *SmCHS* and activate its expression for anthocyanin synthesis. *SmMYB75* was suggested to modulate the anthocyanin's accumulation by combinatorically interacting with a basic helix-loop-helix (*bHLH*) factor. A study analyzing a multiparental population identified two *MYB113* genes on chromosomes 1 and 10 to be involved in anthocyanin accumulation in the fruit (Mangino et al. 2022).

3.3 Glycoalkaloids

Glycoalkaloids are biologically active secondary metabolites, which are chemically defined by the presence of their nitrogenous steroidal aglycone and glycoside residues located at the C-3 position (Zhao et al. 2020; Al Sinani and Eltayeb 2017). The crops of the Solanaceae family are the warehouse of steroidal glycoalkaloids (SGA) which are present in all tissues of eggplant (Sawai et al. 2014) and are better characterized in tomato and potato (Friedman 2002, 2006; Kozukue et al. 2008). The first discovery of SGA is from the species *Solanum nigrum* in 1820. The higher concentration of *Solanum* glycoalkaloids in fruits have a significant impact on the organoleptic properties of flesh by imparting a characteristic bitter taste and off flavor (Aubert et al. 1989a, b; Sánchez-Mata et al. 2010). Eggplant is a source of two main glycoalkaloids namely, water-soluble triglycosides solasonine (SN) and solamargine (SM), which are in general referred to as α-compounds (Trivedi and Pundarikakshudu 2007). The chemical structures of solasonine and solamargine of eggplant are represented in Fig. 3.

The glycoalkaloid accumulation is modulated by several environmental factors during plant growth, harvesting, and post-harvest treatment including light exposure, temperature, altitude, soil type, soil moisture, drought, and soil fertility. Wounding increased glycoalkaloid contents (Sinden et al. 1984; Zhao et al. 2020). Solamargine and solasonine are known to play important roles in plant abiotic and defense stress resistance, assisting plant-environment interactions against a variety of pathogens and predators (Zhao et al. 2020; Al Sinani and Eltayeb 2017). They were also found to have allelopathic effects to suppress seedling growth in crop fields (Vaananen 2007). However, depending on the dose, they can also be used as drugs due to their various biological activities with potential therapeutic potentials (Lee et al. 2007;

Tiossi et al. 2012). But the higher concentration of these glycoalkaloids were known to cause toxic effects on target organisms leading to cell-membrane disruption, acetylcholinesterase inhibition, liver damage, heart damage, teratogenicity, and embryotoxicity (Al Sinani and Eltayeb 2017). Various methods have been employed for analyzing glycoalkaloids in foods and plant materials, of which high performance liquid chromatography (HPLC), high performance thin-layer chromatography (HPTLC), gas chromatography (GC), and mass spectrometric methods and rarely immunoassays are routinely used (Kuronen et al. 1999; Kreft et al. 2000; Dinan et al. 2001; Trivedi and Pundarikakshudu 2007; Vaananen 2007; Eanes et al. 2008).

3.3.1 Biosynthesis Pathway

Cholesterol is a key intermediate for biosynthesis of glycoalkaloids which is obtained as intermediate of acetyl-CoA pathway. SGA biosynthesis begins with the cytosolic mevalonate pathway and a multi-step pathway for the production of cycloartenol, cholesterol, and core SGA, culminating in the synthesis and glycosylation of the steroidal alkaloid aglycone (Zhao et al. 2020). The aglycones are glycosylated through the action of a series of glycosyl transferases (Wang et al. 2017) to form the final products α-solasonine and α-solamargine (Wang et al. 2017) (Fig. 6). The biosynthesis is regulated by a group of genes clustered together in *Solanum* species and the modulation of these genes changes the biosynthesis in the plant (Wang et al. 2017). The site of synthesis of glycoalkaloids has been extensively studied in tomato. The storage of glycoalkaloids occurs in the soluble phase of the cytoplasm and/or in the vacuoles. The alkaloids are not transported throughout the plant system, but they tend to remain confined to the site of synthesis. It is also known that the isopentenyl pyrophosphate (IPP) acts as an intermediate in the biosynthetic pathways of chlorophyll, carotenoid, and glycoalkaloid; therefore, the biosynthesis of each of these three compounds influence that of the others (Zhao et al. 2020). The presence of a gene encoding dioxygenase involved in the

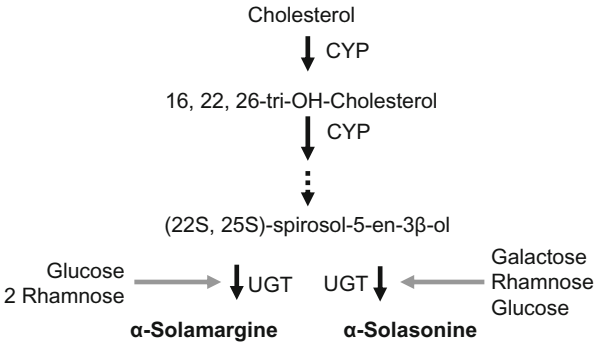


Fig. 6 Suggested biosynthetic pathway for solasonine and solamargine in eggplant. (Adapted from Akiyama et al. (2021)). Enzyme abbreviations are shown in three letters near the arrows: *CYP1* cytochrome *P450* monooxygenase involved in cholesterol hydroxylation, *CYP2* cytochrome *P450* monooxygenase involved in oxygenation, *UGT* uridine diphosphate-dependent glycosyltransferase

conversion of spirosolane alkaloids to solanidane-type alkaloids in eggplant suggest that leaves of eggplant could contain solanidane alkaloids also but have not been tested for its role (Akiyama et al. 2021).

3.3.2 Breeding for Glycoalkaloid Content

Reduction in glycoalkaloid content has been an important breeding objective for crop improvement in eggplant. In certain populations of eggplants studied, the long-fruited cultivars contained a higher concentration of reducing sugars, phenolics, and dry matter than round-fruited types. The range of glycoalkaloid contents in the Indian commercial cultivars varied from 0.37 to 4.83 mg/100 g fresh weight, whereas increase in concentration to 20 mg/100 g fresh weight results in bitter taste and off flavor (Dhruve et al. 2014). The range of glycoalkaloid content in the study of Dhruve et al. (2014) varied from 0.128 to 0.191% with the lowest concentration found in the cultivar 'Doli-5' and significantly higher concentration in 'GOB-1' cultivar. The studies of Sanga et al. (2019) reported the cultivar 'G-3' was found to have the highest concentration of solasodine with the value of 28.25 mg/100 g, whereas lowest concentration of solasodine was in the cultivar 'G-5' (8.78 mg/100 g). Other studies (Lelario et al. 2019) confirmed the phytochemical diversity for alkaloid content in eggplant varieties and others have shown such variation in different species of *Solanum* (Amadi et al. 2013; Eze and Kanu 2014; Chinedu et al. 2011). A significant negative association ($r = -0.588$) was observed between ascorbic acid content and glycoalkaloid content. And glycoalkaloid content had a positive correlation with fruit weight ($r = 0.381$) and fruit volume ($r = 0.399$) (Sanga et al. 2019).

The identification and mapping of genes/QTLs for the contents of solasonine, and solamargine were carried out by Toppino et al. (2016) using an F_2 intraspecific mapping population of 156 individuals ("305E40" × "67/3."). The QTL analysis revealed the presence of one major QTL, *SME06.ML* on chromosome number 6 having the LOD value of 5.02 and PVE of 13.9. The genetic transformation studies paved the path to understand the effects of photosynthetic pigment accumulation on glycoalkaloid biosynthesis in eggplant. Since both chlorophyll and carotenoid biosynthesis pathways share intermediate metabolites with glycoalkaloid synthesis and are regulated by light, it was hypothesized that changes in one pathway could affect the others. Therefore, the study was carried out to silence three key genes involved in carotenoid biosynthesis (PDS) and chlorophyll biosynthesis (*ChlI* and *ChlH*) using TRV-mediated VIGS to block the biosynthesis of respective metabolites. The transformed plants were quantified for carotenoid and chlorophyll levels using liquid chromatography-mass spectrometry. The results revealed a significant reduction in glycoalkaloid production in transformed plants along with the other metabolites in the silenced lines suggesting the crosstalk of both the SGA and chlorophyll biosynthesis. This finding of Toppino et al. (2016) suggests strategies to reduce the levels of endogenous antinutritional compounds in crops. Later, Barchi et al. (2019) studied the evolution of glycoalkaloid biosynthesis in solanaceous crops. In eggplant most core genes for steroidal glycoalkaloid (SGA) biosynthesis genes form two metabolic gene clusters, on chromosome 7 and chromosome 12, while the one on chromosome

12, contains genes named *GLYCOALKALOID METABOLISM 4* and *12* (*GAME4* and *GAME12*).

4 Metabolomics to Identify Novel Eggplant Phenotypes

Metabolomics, where high-throughput identification and quantification of metabolites are done using mass spectral methods, are increasingly applied to study crop plants. Eggplant metabolites have been profiled and attempts have been made to link metabolic profiles to fruit morphology and nutrition, drought stress, nutrient use efficiency, response to pathogens, and other desirable traits (Mauceri et al. 2022; Chen et al. 2021; Hanifah et al. 2018; Mibei et al. 2017).

Hanifah et al. (2018) analyzed eggplant fruit from 21 accessions using untargeted metabolomics with gas chromatography and mass spectrometry in tandem (GC-MS) and liquid chromatography with mass spectrometry (LC-MS) methods. Untargeted metabolomics identified 207 and 51 putative metabolites from GC-MS and LC-MS analysis, respectively, and they belonged to terpenoids, alkaloids, flavonoids, and fatty acid pathways. Two metabolite clusters were identified from GC-MS analysis that further discriminated the 21 eggplant accessions into two respective groups. For example, linoleic acid, palmitic acid, alpha-tocopherol (vitamin E), and neophytadiene metabolites from GS-MS were present across most of the eggplant accessions. Similarly, 2, 5-Bis (N-hexylmethylsilyl) thiophene, 2-(Methylthiomethyl)-3-phenyl-2-propenal, and Boscalid metabolites from LC-MS were present across more than three eggplant accessions. These metabolites represent terpenes and alkaloid pathways. Interestingly, most of the metabolites detected from both GC-MS and LC-MS analysis were accession-specific (Hanifah et al. 2018), suggesting a unique metabolic landscape from fruits of diverse eggplant accessions. Some of these accession-specific metabolites represented the steroids, flavonoids, alkanes, terpenoids, saturated and unsaturated fatty acid pathways. In addition to known metabolites, the LC-MS also detected previously unknown metabolites in a sub-set of eggplant accessions. Furthermore, different fruit characteristics also showed significant correlations with specific metabolic profiles in the 21 eggplant accessions (Hanifah et al. 2018). These fruit traits represented curvature (*CVT*), spininess of calyx (*SCL*), anthocyanin coloration under calyx (*IUC*), depth of indentation of pistil scar (*DPS*), and fruit patches (*PTC*), density of stripes (*DST*), and ribs (*RBS*). For instance, G55 eggplant accession had strong fruit CVT and showed high correlation with 1,2,4-nonadecanetriol. Similarly, PTC trait correlated with farnesyl acetone, and lariciresinol, and clionasterol metabolites, while SCL trait had high correlations with ethyl 9-heptadecenoate and 1-tetradecanoyl-glycero-3-phosphoserine metabolites in different eggplant accessions.

Studies also demonstrated the use of metabolomics to understand the effect of biotic and abiotic stresses in eggplant. For example, eggplant genotypes with high ('AM222') and low ('305E40') nitrogen utilization efficiency (NUE) were characterized for primary and secondary metabolites under nitrogen starvation, and short- and lone-term nitrogen-limiting resupply conditions using GC-MS and

UPLC-qTOF-MS approaches (Mauceri et al. 2022). A differential metabolite analysis revealed several pathways associated with differential NUE in eggplant. For instance, glycine and glyA transcript accumulation in high NUE genotype appears useful for acclimation to the N-limiting stress conditions. An abundance of sucrose and starch pathway metabolites were associated with improved NUE. Similarly, higher levels of L-aspartate and L-asparagine in high NUE genotypes were also identified at short-term low-nitrogen treatment, whereas granule-bound starch synthase and endoglucanase were downregulated under long-term nitrogen stress (Mauceri et al. 2022). These results delineate specific metabolic pathways to target for improving NUE in eggplant. In a similar way, GC-MS was applied to analyze the biochemical changes related to drought stress in African eggplant accessions (Mibei et al. 2017). Nineteen eggplant accessions were analyzed for metabolite changes under regular and water limiting conditions. Drought stress broadly affected the metabolic landscape of eggplant genotypes by changes in sugar, amino acid, and organic acid pathways. The levels of sucrose, fructose, trehalose, xylose, and mannose sugars were positively correlated with the drought stress (Mibei et al. 2017). Likewise, the concentrations of proline, glutamate, citrate, isocitrate, malate, and fumarate were relatively abundant in drought conditions. In contrast, the levels of glucose, myo-inositol, glycine, alanine, quinic acid, aspartate, and malate were less abundant under water-limiting conditions (Mibei et al. 2017). The metabolic changes were defined by genotype, growth stages, and water stress conditions, hence suggesting the complexity of drought stress responses in eggplant.

Recently, a metabolome and volatilome analysis has been conducted to elucidate the biochemical responses of eggplant and tomato against infestation from *Tuta absoluta*, a destructive pest of Solanaceae crops (Chen et al. 2021). A total of 141 volatiles and 797 metabolites were identified that consisted of alcohols, aldehydes, amines, ketones, and terpenes (Chen et al. 2021). The terpene metabolites were differentially enriched in eggplants than tomato. Interestingly, several compounds were either unique in eggplant or were present in lower quantity in tomato in both control and *T. absoluta* infested samples. The study identified about 35 differentially regulated compounds associated with evoked *T. absoluta* response, hence were considered to regulate the pest behavior (Chen et al. 2021). Overall, the differentially accumulated compounds might determine the eggplant resistance against *T. absoluta* infestation, which helps to develop integrated pest control measures for sustainable eggplant production.

5 Eggplant Genomes and Resequencing

Different nuclear genomes of the eggplant have been sequenced within last decade. First draft genome sequence "SME_r2.5.1" of the common eggplant (*Solanum melongena* L.) cultivar 'Nakate-Shinkuro' was assembled using the Roche 454 GS FLX Titanium and Illumina HiSeq 2000 technologies (Hirakawa et al. 2014). It had a total length of 833.1 megabase (Mb) from 33,873 assembled scaffolds. These scaffolds covered about 74% of the predicted 1.12 gigabase (Gb) genome size (Hirakawa

et al. 2014). Annotation of "SME_r2.5.1" genome revealed a total of 85,446 coding sequences, which represented about 48% transposable elements, 2% pseudogenes, and 0.7% short length genes. The repeat sequences constituted about 70% (586.8 Mbs) of the draft genome assembly and mainly represented Gypsy class retrotransposons (Hirakawa et al. 2014). Out of remaining 42,035 genes, 91.6% and 8.4% were designated as "intrinsic" genes having start and stop codons, and partial genes without start and/or stop codons, respectively. A comparative genomic analysis of predicted eggplant genes against other solanaceous plant genomes revealed about 4018 genes exclusive to eggplant genome. In addition, a total of 16,573 orthologous gene pairs were identified between eggplant and tomato that were spread across 56 conserved syntenic blocks. These observations indicated a considerable level of synteny between these two solanaceous crop species.

Due to highly fragmented nature of "SME_r2.5.1" genome, recent efforts from the Eggplant Genome Project (http://www.eggplantgenome.org) led to a chromosome scale fully anchored eggplant genome by integrating single molecule optical mapping and Illumina sequencing approaches (Barchi et al. 2019). This genome assembly from an inbred eggplant line '67/3' consisted of 0.92 Gb ungapped and 1.22 Gb gapped sequences across 469 scaffolds – a significant improvement from the previously available draft genome (Hirakawa et al. 2014). The hybrid scaffolds were anchored to 12 chromosomes using a genetic map of 5964 markers, and the resulting anchored pseudomolecules represented 1.14 Gb gapped and 0.82 Gb ungapped sequences (Barchi et al. 2019). The '67/3' genome had low residual heterozygosity (0.027%) and high (73%) transposable element sequences. It had a total of 34,916 predicted protein-coding genes (Barchi et al. 2019), which were comparable to other Solanaceae genomes (Barchi et al. 2019). A Benchmarking Universal Single-Copy Orthologs (*BUSCO*) score was 96.9%, indicating a near complete gene annotation of the '67/3' eggplant genome. An ortholog analysis with the new eggplant gene models indicated that 667 gene families were specific to eggplant – with a significant enrichment of "pentatricopeptide repeat-containing protein" family than found in tomato, potato, pepper, and Arabidopsis. A survey of single-locus simple sequence repeats (SSRs) identified about 125 SSRs/1 Mbs of '67/3' genome space.

More recently, another chromosome-scale genome of 'GUIQIE-1' eggplant cultivar has been assembled by combining the long read PacBio and Hi-C sequencing approaches (Li et al. 2020). The total length of 'GUIQIE-1' genome is 1155.8 Mb (N50 = 93.9 Mb), out of which, repetitive sequences were 70.1% of the total assembly. Genome annotation identified a total of 35,018 protein-coding genes – a little higher than the annotated genes in the '67/3' eggplant genome. The number of eggplant-specific gene families ($n = 646$) were almost similar between 'GUIQIE-1' and '67/3' genomes. An additional orthologous gene family analysis indicated an expansion in disease resistance genes for bacterial spot resistance in eggplant and pepper than in tomato and potato (Li et al. 2020). Overall, the fully anchored and high-quality genome assemblies from '67/3' and 'GUIQIE-1' accessions provide important resources for genetic improvement and functional genomics studies in eggplant and related species.

With the availability of high-quality chromosome scale genomes, it is now feasible to characterize the genomic diversity and evolutionary footprints of eggplant and its wild relatives. Until now, the genomes of a small set of seven *S. melongena* and one wild *S. incanum* accession has been re-sequenced at 19.8 × depth (Gramazio et al. 2019). Alignments of these eight re-sequenced genomes against '67/3' reference genome identified about nine million single nucleotide polymorphisms (SNPs) and 700,000 Insertion/Deletions (Indels). The wild *S. incanum* accession represented most of this variation. The variants in the *S. melongena* accessions ranged from 0.8 to 1.3 million (Gramazio et al. 2019), indicating a narrow genetic diversity in cultivated eggplant. In addition, about 98% of the variants were present in intergenic and intronic regions. The moderate effect protein-coding (codon insertion/deletion, codon substitution) SNPs ranged from 0.61% to 0.78%, while the large effect SNPs (loss of function mutations or protein truncation) ranged from 0.05% to 0.09% (Gramazio et al. 2019). This analysis provides a preliminary overview of genetic variation and its use for eggplant genetic improvement. Genome sequence resources should allow us to develop and utilize DNA markers for key traits of importance in eggplant that are connected to health promoting metabolites and methods to select desirable plants. Exploration of genomic and phytochemical diversity in the eggplant will also lead to future synthetic biology strategies to overproduce specific phytochemicals of value.

References

Acciarri N, Vitelli G, Arpaia S, Mennella G, Sunseri F et al (2000) Transgenic resistance to the Colorado potato beetle in Bt-expressing eggplant fields. HortScience 35(4):722–725

Ahmed NU, Park JI, Jung HJ, Yang TJ, Hur Y et al (2014) Characterization of dihydroflavonol 4-reductase (DFR) genes and their association with cold and freezing stress in *Brassica rapa*. Gene 550(1):46–55

Akanitapichat P, Phraibung K, Nuchklang K, Prompitakkul S (2010) Antioxidant and hepatoprotective activities of five eggplant varieties. Food Chem Toxicol 48(10):3017–3021

Akiyama R, Watanabe B, Nakayasu M, Lee HJ, Kato J et al (2021) The biosynthetic pathway of potato solanidanes diverged from that of spirosolanes due to evolution of a dioxygenase. Nat Commun 12:1300

Al Sinani SSS, Eltayeb EA (2017) The steroidal glycoalkaloids solamargine and solasonine in *Solanum* plants. S Afr J Bot 112:253–269

Almeida AAP, Farah A, Silva DA, Nunan EA, Glória MBA (2006) Antibacterial activity of coffee extracts and selected coffee chemical compounds against enterobacteria. J Agric Food Chem 54 (23):8738–8743

Amadi B, Onuoha N, Amadi C, Ugbogu A, Duru M (2013) Elemental, amino acid and phytochemical constituents of fruits of three different species of eggplant. Int J Med Arom Plants 3(2):200–203

André CM, Schafleitner R, Legay S, Lefèvre I, Aliaga CAA et al (2009) Gene expression changes related to the production of phenolic compounds in potato tubers grown under drought stress. Phytochemistry 70(9):1107–1116

Astadi IR, Astuti M, Santoso U, Nugraheni PS (2009) In vitro antioxidant activity of anthocyanins of black soybean seed coat in human low density lipoprotein (LDL). Food Chem 112(3):659–663

Aubert S, Daunay MC, Pochard E (1989a) Saponosides steroidiques de l'aubergine (*Solanum melongena* L.). I. Interet alimentaire, methodologie d'analyse, localisation dans le fruit. Agronomie 9(7):641–651

Aubert S, Daunay MC, Pochard E (1989b) Steroidic saponins from egg-plant (*Solanum melongena* L.) II – effects of cultural conditions, genotypes and parthenocarpy. Agronomie 9:751–758

Aubriot X, Knapp S, Syfert MM, Poczai P, Buerki S (2018) Shedding new light on the origin and spread of the brinjal eggplant (*Solanum melongena* L.) and its wild relatives. Am J Bot 105(7): 1175–1118

Aza-Gonzalez C, Herrera-Isidrón L, Núñez-Palenius HG, Martínez De La Vega O et al (2013) Anthocyanin accumulation and expression analysis of biosynthesis-related genes during chili pepper fruit development. Biol Plant 57(1):49–55

Barchi L, Lanteri S, Portis E, Valè G, Volante A et al (2012) A RAD tag derived marker based eggplant linkage map and the location of QTLs determining anthocyanin pigmentation. PLoS One 7:e43740

Barchi L, Pietrella M, Venturini L, Minio A, Toppino L et al (2019) A chromosome-anchored eggplant genome sequence reveals key events in *Solanaceae* evolution. Sci Rep 9:11769

Bhanushree N, Saha P, Tomar BS, Lyngdoh YA, Krishnan SG et al (2019) Genetic analysis and identification of molecular marker linked to the gene for fruit skin colour in eggplant (*Solanum melongena* L.). Veg Sci 45(2):149–153

Borovsky Y, Oren-Shamir M, Ovadia R, De Jong W, Paran I (2004) The A locus that controls anthocyanin accumulation in pepper encodes a MYB transcription factor homologous to Anthocyanin2 of Petunia. Theor Appl Genet 109(1):23–29

Cao G, Sofic E, Prior RL (1996) Antioxidant capacity of tea and common vegetables. J Agric Food Chem 44(11):3426–3431

Casati L, Pagani F, Braga PC, Scalzo RL, Sibilia V (2016) Nasunin, a new player in the field of osteoblast protection against oxidative stress. J Funct Foods 23:474–484

Chalker-Scott L (1999) Environmental significance of anthocyanins in plant stress responses. Photochem Photobiol 70(1):1–9

Cham BE (2012) Intralesion and curaderm BEC5 topical combination therapies of solasodine rhamnosyl glycosides derived from the eggplant or devil's apple result in rapid removal of large skin cancers. Methods of treatment compared. Int J Clin Med 3:115–124

Chami LA, Mendez R, Chataing B, O'Callaghan J, Usubillaga A et al (2003) Toxicological effects of alpha-solamargine in experimental animals. Phytother Res 17:254–258

Chen AY, Chen YC (2013) A review of the dietary flavonoid, kaempferol on human health and cancer chemoprevention. Food Chem 138(4):2099–2107

Chen ZY, Peng C, Jiao R, Wong YM, Yang N et al (2009) Anti-hypertensive nutraceuticals and functional foods. J Agric Food Chem 57(11):4485–4499

Chen L, Li X, Zhang J, He T, Huang J et al (2021) Comprehensive metabolome and volatilome analyses in eggplant and tomato reveal their differential responses to *Tuta absoluta* infestation. Front Plant Sci 12:757230

Chinedu SN, Olasumbo AC, Eboji OK, Emiloju OC, Arinola OK, Dania DI (2011) Proximate and phytochemical analyses of *Solanum aethiopicum* L. and *Solanum macrocarpon* L. fruits. Res J Chem Sci 1(3):63–71

Chioti V, Zeliou K, Bakogianni A, Papaioannou C, Biskinis A, Petropoulos C, Lamari FN, Papasotiropoulos V (2022) Nutritional value of eggplant cultivars and association with sequence variation in genes coding for major phenolics. Plants (Basel) 11(17):2267. https://doi.org/10.3390/plants11172267. PMID: 36079649; PMCID: PMC9460228

Cho AS, Jeon SM, Kim MJ, Yeo J, Seo KI et al (2010) Chlorogenic acid exhibits anti-obesity property and improves lipid metabolism in high-fat diet-induced-obese mice. Food Chem Toxicol 48(3):937–943

Chumyam A, Whangchai K, Jungklang J, Faiyue B, Saengnil KJSA (2013) Effects of heat treatments on antioxidant capacity and total phenolic content of four cultivars of purple skin eggplants. Sci Asia 39:246–251

Da Costa GAF, Morais MG, Saldanha AA, Assis Silva IC, Aleixo ÁA et al (2015) Antioxidant, antibacterial, cytotoxic, and anti-inflammatory potential of the leaves of *Solanum lycocarpum* A. St. Hil. (Solanaceae). Evid Based Complement Alternat Med 2015:315987

Daunay MC (2008) Eggplant. In: Prohens J, Nuez F (eds) Vegetables II. Springer, New York, pp 163–220

Daunay MC, Hazra P (2012) Eggplant. In: Peter KV, Hazra P (eds) Handbook of vegetables. Studium Press, Houston, pp 257–322

Daunay MC, Lester RN, Gebhardt C, Hennart JW, Jahn M (2001) Genetic resources of eggplant (*Solanum melongena* L.) and allied species: a new challenge for molecular geneticists and eggplant breeders. In: van den Berg RG, Barendse GWM, van der Weerdem GM, Mariani C (eds) Solanaceae V. Advances in taxonomy and utilization. Nijmegen University Press, Nijmegen, pp 251–274

DeFuria J, Bennett G, Strissel KJ, Perfield JW, Milbury PE et al (2009) Dietary blueberry attenuates whole-body insulin resistance in high fat-fed mice by reducing adipocyte death and its inflammatory sequelae. J Nutr 139(8):1510–1516

Dhruve JJ, Shah R, Gandhi S, Talati J (2014) Biochemical and morphological traits of different cultivars of brinjal fruits growing in Anand (Gujarat). Indian J Agric Biochem 27(2):211–214

Di Guardo M, Tadiello A, Farneti B, Lorenz G, Masuero D et al (2013) A multidisciplinary approach providing new insight into fruit flesh browning physiology in apple (*Malus x domestica* Borkh.). PLoS One 8(10):e78004

Dinan L, Harmatha J, Lafont R (2001) Chromatographic procedures for the isolation of plant steroids. J Chromatogr A 935(1–2):105–123

Docimo T, Francese G, De Palma M, Mennella D, Toppino L et al (2016a) Insights in the fruit flesh browning mechanisms in *Solanum melongena* genetic lines with opposite postcut behavior. J Agric Food Chem 64(22):4675–4685

Docimo T, Francese G, Ruggiero A, Batelli G, De Palma M et al (2016b) Phenylpropanoids accumulation in eggplant fruit: characterization of biosynthetic genes and regulation by a MYB transcription factor. Front Plant Sci 6:1233

Donmez C, Yallc FN, Boyacioglu O, Korkussuz P, Akkol EK et al (2020) From nutrition to medicine: assessing hemorrhoid healing activity of *Solanum melongena* L. via in vivo experimental models and its major chemicals. J Ethanopharmacol 261:1131–1143

Donzella G, Spena A, Rotino GL (2000) Transgenic parthenocarpic eggplants: superior germplasm for increased winter production. Mol Breed 6(1):79–86

Dos Santos MD, Almeida MC, Lopes NP, De Souza GEP (2006) Evaluation of the anti-inflammatory, analgesic and antipyretic activities of the natural polyphenol chlorogenic acid. Biol Pharm Bull 29(11):2236–2240

Eanes RC, Tek N, Kirsoy O, Frary A, Doganlar S, Almeida AE (2008) Development of practical HPLC methods for the separation and determination of eggplant steroidal glycoalkaloids and their aglycones. J Liquid Chromatogr Relat Technol 31(7):984–1000

El-Seedi HR, El-Said AMA, Khalifa SAM, Görensson U, Bohlin L et al (2012) Biosynthesis, natural sources, dietary intake, pharmacokinetic properties, and biological activities of hydroxycinnamic acids. J Agric Food Chem 60(44):10877–10895

Eze SO, Kanu CQ (2014) Phytochemical and nutritive composition analysis of *Solanum aethopicum* L. J Pharm Sci Innov 3(4):358–362

Florio FE, Gattolin S, Toppino L, Bassolino L, Fibiani M et al (2021) A SmelAAT acyltransferase variant causes a major difference in eggplant (*Solanum melongena* L.) peel anthocyanin composition. Int J Mol Sci 22:9174. https://doi.org/10.3390/ijms22179174

Fraikue FB (2016) Unveiling the potential utility of eggplant: a review. In: Conference proceedings of international conference on education, development and innovation, Tang Palace Hotel, Accra, pp 883–895

Frary A, Doganlar S, Daunay MC (2007) Eggplant. In: Kole C (ed) Genome mapping and molecular breeding in plants. Volume 5: vegetables. Springer, Berlin/Heidelberg, pp 287–313

Friedman M (2002) Tomato glycoalkaloids: role in the plant and in the diet. J Agric Food Chem 50(21):5751–5780

Friedman M (2006) Potato glycoalkaloids and metabolites: roles in the plant and in the diet. J Agric Food Chem 54(23):8655–8681

Gajewski M, Katarzyna K, Bajer M (2009) The influence of postharvest storage on quality characteristics of fruit of eggplant cultivars. Not Bot Hortic Agrobot Cluj Napoca 37(2):200–205

García-Salas P, Gómez-Caravaca AM, Morales-Soto A, Segura-Carretero A, Fernández-Gutiérrez A (2014) Identification and quantification of phenolic compounds in diverse cultivars of

eggplant grown in different seasons by high-performance liquid chromatography coupled to diode array detector and electrospray-quadrupole-time of flight-mass spectrometry. Food Res Int 57:114–122
Gramazio P, Prohens J, Plazas M, Andújar I, Herraiz FJ et al (2014) Location of chlorogenic acid biosynthesis pathway and polyphenol oxidase genes in a new interspecific anchored linkage map of eggplant. BMC Plant Biol 14(1):1–15
Gramazio P, Yan H, Hasing T, Vilanova S, Prohens J et al (2019) Whole-genome resequencing of seven eggplant (*Solanum melongena*) and one wild relative (*S. incanum*) accessions provides new insights and breeding tools for eggplant enhancement. Front Plant Sci 10:1220
Grubben GJH, Tindall HD, Williams JT (1977) Eggplant. In: Tindall HD, Williams JT (eds) Tropical vegetables and their genetic resources. IBPGR, Rome, pp 34–37
Guimaraes PR, Glavao AM, Batista CM, Azevedo GS, Oliveira RD et al (2000) Eggplant (*Solanum melongena*) infusion has a modest and transitory effect on hypercholesterolemic subjects. Braz J Med Biol Res 33:1027–1036
Guo J, Han W, Wang M (2008) Ultraviolet and environmental stresses involved in the induction and regulation of anthocyanin biosynthesis: a review. Afr J Biotechnol 7(25):4966–4972
Hallmann E (2012) The influence of organic and conventional cultivation systems on the nutritional value and content of bioactive compounds in selected tomato types. J Sci Food Agric 92(14): 2840–2848
Hallmann E, Rembiałkowska E (2012) Characterisation of antioxidant compounds in sweet bell pepper (*Capsicum annuum* L.) under organic and conventional growing systems. J Sci Food Agric 92(12):2409–2415
Hanifah A, Maharijaya A, Putri SP, Lavina WA, Sobir (2018) Untargeted metabolomics analysis of eggplant (*Solanum melongena* L.) fruit and its correlation to fruit morphologies. Metabolites 8:49
Hanson PM, Yang RY, Tsou SC, Ledesma D, Engle L et al (2006) Diversity in eggplant (*Solanum melongena*) for superoxide scavenging activity, total phenolics, and ascorbic acid. J Food Compos Anal 19(6–7):594–600
Harisha SM, Saha P, Lyngdoh YA, Tomar BS, Kundu A (2023) Screening of eggplant genotypes with respect to anthocyanin content. Indian J Hortic 80(1):44–49
Haytowitz DB, Ahuja JKC, Wu X, Somanchi M, Nickle M et al (2019) USDA National Nutrient Database for Standard Reference, Legacy Release. Nutrient Data Laboratory, Beltsville Human Nutrition Research Center, ARS, USDA. https://data.nal.usda.gov/dataset/usda-national-nutrient-database-standard-reference-legacy-release. Accessed 5 Sept 2022
He J, Giusti MM (2010) Anthocyanins: natural colorants with health-promoting properties. Annu Rev Food Sci Technol 1:163–187
Helmja K, Vaher M, Gorbatšova J, Kaljurand M (2007) Characterization of bioactive compounds contained in vegetables of the *Solanaceae* family by capillary electrophoresis. Proc Estonian Acad Sci 56(4):172–186
Hirakawa H, Shirasawa K, Miyatake K, Nunome T, Negoro S et al (2014) Draft genome sequence of eggplant (*Solanum melongena* L.): the representative *Solanum* species indigenous to the old world. DNA Res 21:649–660
Jaakola L (2013) New insights into the regulation of anthocyanin biosynthesis in fruits. Trends Plant Sci 18(9):477–483
Jan AT, Kamli MR, Murtaza I, Singh JB, Ali A (2010) Dietary flavonoid quercetin and associated health benefits – an overview. Food Rev Int 26(3):302–317
Jaramillo S, Lopez S, Verale LM, Arcos RR, Jimenez A et al (2010) The flavonol isorhamnetin exhibits cytotoxic effects on human colon cancer cells. J Agric Food Chem 58(20):10869–10875
Jiang D, Li D, Wu W (2013) Inhibitory effects and mechanisms of luteolin on proliferation and migration of vascular smooth muscle cells. Nutrients 5(5):1648–1659
Jiang M, Liu Y, Ren L, Lian H, Chen H (2016) Molecular cloning and characterization of anthocyanin biosynthesis genes in eggplant (*Solanum melongena* L.). Acta Physiol Plant 38(7):1–13
Jin XH, Ohgami K, Shiratori K, Suzuki Y, Koyama Y et al (2006) Effects of blue honeysuckle (*Lonicera caerulea* L.) extract on lipopolysaccharide-induced inflammation in vitro and in vivo. Exp Eye Res 82(5):860–867

Kalalinia F, Karimi-Sani I (2017) Anticancer properties of solamargine: a systematic review. Phytother Res 31:858–870

Kashyap V, Kumar SV, Collonnier C, Fusari F, Haicour R et al (2003) Biotechnology of eggplant. Sci Hortic 97(1):1–25

Kaur C, Nagal S, Nishad J, Kumar R (2014) Evaluating eggplant (*Solanum melongena* L.) genotypes for bioactive properties: a chemometric approach. Food Res Int 60:205–211

Kaushik P, Gramazio P, Vilanova S, Raigón MD, Prohens J et al (2017) Phenolics content, fruit flesh colour and browning in cultivated eggplant, wild relatives and interspecific hybrids and implications for fruit quality breeding. Int Food Res J 102:392–401

Kaushik P, Kumar P, Kumar S (2020) Enhancement of chlorogenic content of the eggplant fruit with eggplant hydroxycinnamoyl CoA-quinate transferase gene via novel agroinfiltration protocol. Pharmacogn Mag 16(5):450

Knapp S, Vorontsova MS, Prohens J (2013) Wild relatives of the eggplant (*Solanum melongena* L.; *Solanaceae*): new understanding of species in a complex group. PLoS One 8(2):e57039

Koley TK, Tiwari SK, Sarkar A, Nishad J, Goswami A et al (2019) Antioxidant potential of Indian eggplant: comparison among white, purple and green genotypes using chemometrics. Agric Res 8(1):9–20

Kong JM, Chia LS, Goh NK, Chia TF, Brouillard R (2003) Analysis and biological activities of anthocyanins. Phytochemistry 64(5):923–933

Kozukue N, Yoon KS, Byun GI, Misoo S, Levin CE et al (2008) Distribution of glycoalkaloids in potato tubers of 59 accessions of two wild and five cultivated *Solanum* species. J Agric Food Chem 56(24):11920–11928

Kreft S, Zel J, Pukl M, Umek A, Strukelj B (2000) Non-aqueous capillary electrophoresis for the simultaneous analysis of solasodine and solasonine. Phytochem Anal 11(1):37–40

Kritchevsky D, Tepper SA, Story JA (1975) Influence of an eggplant (*Solanum melongena*) preparation on cholesterol metabolism in rates. Exp Pathol 10:180–183

Kumar R, Munshi AD, Saha P, Behera TK, Lyngdoh YA, Tomar BS (2020) Pusa Hara Baingan 1: A new green brinjal for commercial cultivation. Indian Horticulture 65(4). Retrieved from https://epubs.icar.org.in/index.php/IndHort/article/view/107287

Kumari M, Jain S (2012) Tannin: an antinutrient with positive effect to manage diabetes. Res J Recent Sci 1(12):1–8

Kuronen P, Väänänen T, Pehu E (1999) Reversed-phase liquid chromatographic separation and simultaneous profiling of steroidal glycoalkaloids and their aglycones. J Chromatogr A 863(1):25–35

Kwon YI, Apostolidis E, Shetty KJBT (2008) In vitro studies of eggplant (*Solanum melongena*) phenolics as inhibitors of key enzymes relevant for type 2 diabetes and hypertension. Bioresour Technol 99(8):2981–2988

Lee MH, Cheng JJ, Lin CY, Chen YJ et al (2007) Precursor-feeding strategy for the production of solanine, solanidine and solasodine by a cell culture of *Solanum lyratum*. Process Biochem 42(5):899–903

Lelario F, De Maria S, Rivelli AR, Russo D, Milella L, Bufo SA, Scrano L (2019) A complete survey of glycoalkaloids using LC-FTICR-MS and IRMPD in a commercial variety and a local landrace of eggplant (Solanum melongena L.) and their anticholinesterase and antioxidant activities. Toxins 11(4):230. https://doi.org/10.3390/toxins11040230

Li Y, Ding Y (2012) Mini review: therapeutic potential of myricetin in diabetes mellitus. Food Sci Hum Wellness 1(1):19–25

Li J, Ren L, Gao Z, Jiang M, Liu Y et al (2017) Combined transcriptomic and proteomic analysis constructs a new model for light-induced anthocyanin biosynthesis in eggplant (*Solanum melongena* L.). Plant Cell Environ 40:3069–3087

Li D, Qian J, Li W, Yu N, Gan G et al (2020) A high-quality genome assembly of the eggplant provides insights into the molecular basis of disease resistance and chlorogenic acid synthesis. Mol Ecol Resour 21:1274–1286

Lin W, Li Z (2017) Blueberries inhibit cyclooxygenase-1 and cyclooxygenase-2 activity in human epithelial ovarian cancer. Oncol Lett 13(6):4897–4904

Lo Scalzo R, Fibiani M, Mennella G, Rotino GL, Dal Sasso M et al (2010) Thermal treatment of eggplant (*Solanum melongena* L.) increases the antioxidant content and the inhibitory effect on human neutrophil burst. J Agric Food Chem 58(6):3371–3379

Lorenzo Y, Azqueta A, Luna L, Bonilla F, Dominguez G et al (2009) The carotenoid beta-cryptoxanthin stimulates the repair of DNA oxidation damage in addition to acting as an antioxidant in human cells. Carcinogenesis 30(2):308–314

Løvdal T, Olsen KM, Slimestad R, Verheul M, Lillo C (2010) Synergetic effects of nitrogen depletion, temperature, and light on the content of phenolic compounds and gene expression in leaves of tomato. Phytochemistry 71:605–613

Luthria D, Singh AP, Wilson T, Vorsa N, Banuelos GS et al (2010) Influence of conventional and organic agricultural practices on the phenolic content in eggplant pulp: plant-to-plant variation. Food Chem 121(2):406–411

Lutz M, Henríquez C, Escobar M (2011) Chemical composition and antioxidant properties of mature and baby artichokes (*Cynara scolymus* L.), raw and cooked. J Food Compos Anal 24(1): 49–54

Magania AA, Kamimura N, Soumyanath A, Stevens JF, Maier CS (2021) Caffoylquinic acids: chemistry, biosynthesis, occurrence, analytical challenges, and bioactivity. Plant J 107:1299–1319

Maier A, Hoecker U (2015) COP1/SPA ubiquitin ligase complexes repress anthocyanin accumulation under low light and high light conditions. Plant Signal Behav 10(1):e970440

Mangino G, Arrones A, Plazas M, Pook T, Prohens J, Gramazio P, Vilanova S (2022) Newly developed MAGIC population allows identification of strong associations and candidate genes for anthocyanin pigmentation in eggplant. bioRxiv 2021.09.10.459758; https://doi.org/10.1101/2021.09.10.459758

Manikandan R, Thiagarajan R, Goutham G, Arumugam M, Beulaja M et al (2016) Zeaxanthin and ocular health, from bench to bedside. Fitoterapia 109:58–66

Matsubara K, Kaneyuki T, Miyake T, Mori M (2005) Antiangiogenic activity of nasunin, an antioxidant anthocyanin, in eggplant peels. J Agric Food Chem 53(16):6272–6275

Matsui T, Ebuchi S, Fukui K, Matsugano K, Terahara N et al (2004) Caffeoylsophorose, a new natural α-glucosidase inhibitor, from red vinegar by fermented purple-fleshed sweet potato. Biosci Biotechnol Biochem 68(11):2239–2246

Matsumaru K, Kamihama T, Inada K (1971) Effect of covering materials with different transmission properties on anthocyanin content of eggplant pericarp. Environ Control Biol 9(3–4):9–15

Mauceri A, Aci MM, Toppino L, Panda S, Meir S et al (2022) Uncovering pathways highly correlated to NUE through a combined metabolomics and transcriptomics approach in eggplant. Plants 11:700

Mazza G (2007) Anthocyanins and heart health. Ann Ist Super Sanita 43(4):369

Mei X, Yi C, Huang G (2017) The antioxidant activities of polysaccharides and their derivatives. Curr Drug Targets 18(11):1296–1300

Mennella G, Rotino GL, Fibiani M, D'Alessandro A, Francese G et al (2010) Characterization of health-related compounds in eggplant (*Solanum melongena* L.) lines derived from introgression of allied species. J Agric Food Chem 58(13):7597–7603

Mennella G, Lo Scalzo R, Fibiani M, D'Alessandro A, Francese G et al (2012) Chemical and bioactive quality traits during fruit ripening in eggplant (*S. melongena* L.) and allied species. J Agric Food Chem 60(47):11821–11831

Meyer RS, Karol KG, LittleDP NMH, Litt A (2012) Phylogeographic relationships among Asian eggplants and new perspectives on eggplant domestication. Mol Phylogenet Evol 63(3):685–701

Meyer RS, Bamshad M, Fuller DQ, Litt A (2014) Comparing medicinal uses of eggplant and related *Solanaceae* in China, India, and the Philippines suggests the independent development of uses, cultural diffusion, and recent species substitutions. Econ Bot 68(2):137–152

Mibei EK, Owino WO, Ambuko J, Giovannoni JJ, Onyango AN (2017) Metabolomic analyses to evaluate the effect of drought stress on selected African eggplant accessions. J Sci Food Agric 98:205–216

Mishra BB, Gautam S, Sharma A (2013) Browning of fresh-cut eggplant: impact of cutting and storage. Postharvest Biol Technol 67:44–51

Morishita H, Ohnishi M (2001) Absorption, metabolism and biological activities of chlorogenic acids and related compounds. Stud Nat Prod Chem 25:919–953

Nandi L, Saha P, Jaiswal S, Lyngdoh YA, Behera TK, Pan RS, Munshi AD, Saha ND, Hossain F, Srivastava M, Tomar BS (2021) Bioactive compounds, antioxidant activity and elements content variation in indigenous and exotic *Solanum* sp. and their suitability in recommended daily diet. Sci Hortic 287:110232. https://doi.org/10.1016/j.scienta.2021.110232

Nayak B, Liu RH, Tang J (2015) Effect of processing on phenolic antioxidants of fruits, vegetables, and grains – a review. Crit Rev Food Sci Nutr 55(7):887–918

Newman SM, Eannetta NT, Yu H, Prince JP, Carmen de Vicente M et al (1993) Organisation of the tomato polyphenol oxidase gene family. Plant Mol Biol 21(6):1035–1051

Niño-Medina G, Urías-Orona V, Muy-Rangel MD, Heredia JB (2017) Structure and content of phenolics in eggplant (*Solanum melongena*) – a review. S Afr J Bot 111:161–169

Nisha P, Nazar PA, Jayamurthy P (2009) A comparative study on antioxidant activities of different varieties of *Solanum melongena*. Food Chem Toxicol 47(10):2640–2644

Okmen B, Sigva HO, Mutlu S, Doganlar S, Yemenicioglu A et al (2009) Total antioxidant activity and total phenolic contents in different Turkish eggplant (*Solanum melongena* L.) cultivars. Int J Food Prop 12(3):616–624

Pal JK, Singh M, Rai M, Satpathy S, Singh DV et al (2009) Development and bioassay of Cry1Ac-transgenic eggplant (*Solanum melongena* L.) resistant to shoot and fruit borer. J Hortic Sci Biotechnol 84(4):434–438

Plazas M, Andujar I, Vilanova S, Hurtado M, Gramazio P et al (2013a) Breeding for chlorogenic acid content in eggplant: interest and prospects. Not Bot Hortic Agrobot Cluj Napoca 41(1):26–35

Plazas M, López-Gresa MP, Vilanova S, Torres C, Hurtado M et al (2013b) Diversity and relationships in key traits for functional and apparent quality in a collection of eggplant: fruit phenolics content, antioxidant activity, polyphenol oxidase activity, and browning. J Agric Food Chem 61(37):8871–8879

Pojer E, Mattivi F, Johnson D, Stockley CS (2013) The case for anthocyanin consumption to promote human health: a review. Compr Rev Food Sci Food Saf 12(5):483–508

Povero G, Gonzali S, Bassolino L, Mazzucato A, Perata P (2011) Transcriptional analysis in high-anthocyanin tomatoes reveals synergistic effect of Aft and atv genes. J Plant Physiol 168(3): 270–279

Praca JM, Thomaz A, Caramelli B (2004) Eggplant (*Solanum melongena*) extract does not alter serum lipid levels. Arq Bras Cardiol 82:269–276

Prohens J, Rodríguez-Burruezo A, Raigón MD, Nuez F (2007) Total phenolic concentration and browning susceptibility in a collection of different varietal types and hybrids of eggplant: implications for breeding for higher nutritional quality and reduced browning. J Am Soc Hortic Sci 132(5):638–646

Prohens J, Whitaker BD, Plazas M, Vilanova S, Hurtado M et al (2013) Genetic diversity in morphological characters and phenolic acids content resulting from an interspecific cross between eggplant, *Solanum melongena*, and its wild ancestor (*S. incanum*). Ann Appl Biol 162(2):242–257

Raigón MD, Prohens J, Muñoz-Falcón JE, Nuez F (2008) Comparison of eggplant landraces and commercial varieties for fruit content of phenolics, minerals, dry matter and protein. J Food Compos Anal 21(5):370–376

Raigon MD, Rodriguez-Burruezo A, Prohens J (2010) Effects of organic and conventional cultivation methods on composition of eggplant fruits. J Agric Food Chem 58(11):6833–6840

Ramírez EC, Virador VM (2002) Polyphenol oxidase. In: Handbook of food enzymology. Marcel Dekker, New York, pp 509–523

Ranil RHG, Prohens I, Aubriot X, Niran HML, Plazas M et al (2017) *Solanum insanum* L. (subgenus *Leptostemonum Bitter*, *Solanaceae*), the neglected wild progenitor of eggplant (*S. melongena* L.): a review of taxonomy, characteristics and uses aimed at its enhancement for improved eggplant breeding. Genet Resour Crop Evol 64:1707–1722

Raybould A, Poppy GM (2012) Commercializing genetically modified crops under EU regulations: objectives and barriers. GM Crops Food 3(1):9–20

Rosa-Martínez E, García-Martínez MD, Adalid-Martínez AM, Pereira-Dias L, Casanova C, Soler E et al (2021) Fruit composition profile of pepper, tomato and eggplant varieties grown under uniform conditions. Food Res Int 147:110531

Rotino GL, Sala T, Toppino L (2014) Eggplant. In: Pratap A, Kumar J (eds) Alien gene transfer in crop plants, vol 2. Springer, New York, pp 381–409

Saha P, Lyngdoh YA, Tomar BS (2016) Screening of cultivated eggplant (*Solanum melongena* L.) and wild *Solanum* sp. against Fusarium wilt in India. Indian Phytopathol 69(4s):729–733

Saha P, Munshi AD, Behera TK, Tomar BS, Lyngdoh YA (2021) Pusa Oishiki: New high yielding variety of brinjal. Indian Horticulture 65(6). Retrieved from https://epubs.icar.org.in/index.php/IndHort/article/view/111469

Saha P, Tomar BS, Munshi AD, Behera TK, Lyngdoh YA, Rana S (2022) Pusa Vaibhav: A new round purple variety of brinjal for commercial cultivation. Indian Horticulture 67(4). Retrieved from https://epubs.icar.org.in/index.php/IndHort/article/view/115685

Saha P, Tomar BS, Munshi AD, Behera TK, Kumar R, Lyngdoh YA, Rana SC, Das SK, Rahman FH, Roy B, Sultana S (2023) Brinjal hybrid Pusa Unnat – A hope for higher income for the farmers of Terai region of West Bengal. Indian Horticulture 68(1):50. Retrieved from https://epubs.icar.org.in/index.php/IndHort/article/view/117531

Salehi B, Sharifi-Rad J, Cappellini F, Reiner Ž, Zorzan D et al (2020) The therapeutic potential of anthocyanins: current approaches based on their molecular mechanism of action. Front Pharmacol 11:1300

Sánchez-Mata MC, Yokoyama WE, Hong YJ, Prohens J (2010) α-Solasonine and α-solamargine contents of gboma (*Solanum macrocarpon* L.) and scarlet (*Solanum aethiopicum* L.) eggplants. J Agric Food Chem 58(9):5502–5508

Sanga L, Pandey AK, Sarma P, Singh S (2019) Studies on biochemical profile of wild brinjal (*Solanum gilo* Raddi). Journal of Pharmacognosy and Phytochemistry 8:1138–1140

Särkinen T, Bohs L, Olmstead RG, Knapp S (2013) A phylogenetic framework for evolutionary study of the nightshades (*Solanaceae*): a dated 1000-tip tree. BMC Evol Biol 13(1):1–15

Sato Y, Itagaki S, Kurokawa T, Ogura J, Kobayashi M et al (2011) In vitro and in vivo antioxidant properties of chlorogenic acid and caffeic acid. Int J Pharm 403(1–2):136–138

Sawai S, Ohyama K, Yasumoto S, Seki H, Sakuma T et al (2014) Sterol side chain reductase 2 is a key enzyme in the biosynthesis of cholesterol, the common precursor of toxic steroidal glycoalkaloids in potato. Plant Cell 26(9):3763–3774

Scalzo RL, Fibiani M, Francese G, D'Alessandro A, Rotino GL et al (2016) Cooking influence on physico-chemical fruit characteristics of eggplant (*Solanum melongena* L.). Food Chem 194: 835–842

Shen KH, Hung JH, Chang CW, Weng YT, Wu MJ et al (2017) Solasodine inhibits invasion of human lung cancer cell through downregulation of miR-21 and MMPs expression. Chem Biol Interact 268:129–135

Shetty SM, Chandrashekar A, Venkatesh YP (2011) Eggplant polyphenol oxidase multigene family: cloning, phylogeny, expression analyses and immunolocalization in response to wounding. Phytochemistry 72(18):2275–2287

Sheu MJ, Chou PY, Cheng HC, Wu CH, Huang GJ et al (2009) Analgesic and anti-inflammatory activities of a water extract of *Trachelospermum jasminoides* (*Apocynaceae*). J Ethnopharmacol 126(2):332–338

Shi S, Liu Y, He Y, Li L, Li D et al (2021) R2R3-MYB transcription factor SmMYB75 promotes anthocyanin biosynthesis in eggplant (*Solanum melongena* L.). Sci Hortic 282:110020

Sinden SL, Sanford LL, Webb RE (1984) Genetic and environmental control of potato glycoalkaloids. Am Potato J 61(3):141–156

Singh AP, Luthria D, Wilson T, Vorsa N, Singh V et al (2009) Polyphenols content and antioxidant capacity of eggplant pulp. Food Chem 114(3):955–961

Solberg S, van Zonneveld M, Rakha MT, Taher DI, Prohens J, Jarret R, van Dooijeweert W, Peter Giovannini (2022) Global strategy for the conservation and use of eggplants. Global Crop Diversity Trust. Bonn, Germany

Stommel JR, Whitaker BD (2003) Phenolic acid content and composition of eggplant fruit in a germplasm core subset. J Am Soc Hortic Sci 128(5):704–710
Sun T, Simon PW, Tanumihardjo SA (2009) Antioxidant phytochemicals and antioxidant capacity of biofortified carrots (*Daucus carota* L.) of various colors. J Agric Food Chem 57(10):4142–4147
Swarup V (1995) Genetic resources and breeding of aubergine (*Solanum melongena* L.). Acta Hortic 412:71–79
Taher D, Solberg SO, Prohens J, Chou YY, Rakha M et al (2017) World vegetable center eggplant collection: origin, composition, seed dissemination and utilization in breeding. Front Plant Sci 8:1484
Thipyapong P, Stout MJ, Attajarusit J (2007) Functional analysis of polyphenol oxidases by antisense/sense technology. Molecules 12(8):1569–1595
Tiossi RFJ, Miranda MA, de Sousa JPB, Praça FSG, Bentley MVLB et al (2012) A validated reverse phase HPLC analytical method for quantitation of glycoalkaloids in *Solanum lycocarpum* and its extracts. J Anal Methods Chem 2012(947836). https://doi.org/10.1155/2012/947836
Toppino L, Barchi L, Lo Scalzo R, Palazzolo E, Francese G (2016) Mapping quantitative trait loci affecting biochemical and morphological fruit properties in eggplant (*Solanum melongena* L.). Front Plant Sci 7:256
Tran LT, Taylor JS, Constabel CP (2012) The polyphenol oxidase gene family in land plants: lineage-specific duplication and expansion. BMC Genomics 13(1):1–12
Trivedi P, Pundarikakshudu K (2007) Novel TLC densitometric method for quantification of solasodine in various *Solanum* species, market samples and formulations. Chromatographia 65(3):239–243
Urbany C, Stich B, Schmidt L, Simon L, Berding H et al (2011) Association genetics in *Solanum tuberosum* provides new insights into potato tuber bruising and enzymatic tissue discoloration. BMC Genomics 12(1):1–14
Vaananen T (2007) Glycoalkaloid content and starch structure in *Solanum* species and interspecific somatic potato hybrids (dissertation). EKT-series 1384. Department of Applied Chemistry and Microbiology, University of Helsinki, Helsinki, pp 79 + 45
van Lent DM, Leermakers ETM, Darweesh SKL, Moreira EM, Tielemans MJ et al (2016) The effects of lutein on respiratory health across the life course: a systematic review. Clin Nutr ESPEN 13:e1–e7
Wang LS, Stoner GD (2008) Anthocyanins and their role in cancer prevention. Cancer Lett 269(2): 281–290
Wang CC, Sulli M, Fu DQ (2017) The role of phytochromes in regulating biosynthesis of sterol glycoalkaloid in eggplant leaves. PLoS One 12(12):e0189481
Watson RR, Schönlau F (2015) Nutraceutical and antioxidant effects of a delphinidin-rich maqui berry extract Delphinol: a review. Minerva Cardioangiol 63(2):1–12
Weese T, Bohs L (2010) Eggplant origins: out of Africa, into the Orient. Taxon 59(1):49–56
Whitaker BD, Stommel JR (2003) Distribution of hydroxycinnamic acid conjugates in fruit of commercial eggplant (*Solanum melongena* L.) cultivars. J Agric Food Chem 51(11):3448–3454
Wu F, Eannetta NT, Xu Y, Tanksley SD (2009) A detailed synteny map of the eggplant genome based on conserved ortholog set II (COSII) markers. Theor Appl Genet 118(5):927–935
Xu W, Dubos C, Lepiniec L (2015) Transcriptional control of flavonoid biosynthesis by MYB–bHLH–WDR complexes. Trends Plant Sci 20(3):176–185
Zhang Y, Chu G, Hu Z, Gao Q, Cui B et al (2016) Genetically engineered anthocyanin pathway for high health-promoting pigment production in eggplant. Mol Breed 36(5):1–14
Zhao Y, Wang J, Ballevre O, Luo H, Zhang W (2012) Antihypertensive effects and mechanisms of chlorogenic acids. Hypertens Res 35(4):370–374
Zhao DK, Zhao Y, Chen SY, Kennelly EJ (2020) *Solanum* steroidal glycoalkaloids: structural diversity, biological activities, and biosynthesis. Nat Prod Rep 38(8):1423–1444
Zhao J, Dan Y, Liu Z, Wang Q, Jiang M et al (2022) Solamargine alleviated UVB-induced inflammation and melanogenesis in human keratinocytes and melanocytes via the p38 MAPK signaling pathway, a promising agent for post-inflammatory hyperpigmentation. Front Med 13:812653

Genome Designing for Nutritional Quality in *Amaranthus*

Isadora Louise Alves da Costa Ribeiro Quintans, Valesca Pandolfi, Thais Gaudencio do Rêgo, José Ribamar Costa Ferreira Neto, Thais A. R. Ramos, and Dinesh Adhikary

Contents

I. L. A. da Costa Ribeiro Quintans (✉)
Universidade Federal Rural do Semi-Árido, Mossoro, Brazil
e-mail: isadoralouise@ufersa.edu.br

V. Pandolfi · J. R. C. F. Neto
Universidade Federal de Pernambuco, Recife, Brazil
e-mail: valesca.pandolfi@ufpe.br; netocostaferreira@gmail.com

T. Gaudencio do Rêgo · T. A.R. Ramos
Universidade Federal da Paraíba, João Pessoa, Brazil
e-mail: gaudenciothais@gmail.com; thaisratis@gmail.com

D. Adhikary
Department of Agricultural, Food & Nutritional Sciences, University of Alberta, Edmonton, AB, Canada
e-mail: dadhika1@ualberta.ca

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_56

Abstract

Known for its high nutritional value and ability to thrive in the most adverse conditions, amaranth grain is considered a superfood by many. Amaranth is gluten-free and rich in protein, especially lysine, an essential amino acid very scarce in other plants. The antioxidant properties of amaranth may improve the nutritional status and health of consumers. In fact, as an antioxidant, anti-inflammatory, and anticancer agent, amaranth has been pharmacologically reported to be highly effective against chronic inflammation. Other benefits of amaranth include lowering cholesterol and preventing cardiovascular disease. Even amaranth antinutrients, such as phenolic compounds, have antioxidant properties. Amaranth can be used for the biofortification of staple food by simply adding amaranth compounds to flour or bakery products. A more complex approach involves creating transgenic plants with higher nutritional profiles. Amaranth biofortification may also increase yield and resistance to multiple diseases. Despite these unique characteristics, amaranth is considered an underutilized grain, with greater potential to overcome hunger and diseases. One of the limitations of amaranth cultivation is the lack of high-yielding varieties. Therefore, we provide in this chapter a broad overview of amaranth grain cultivation and the current state of the art of genetic breeding of this crop. We highlight plant breeding using molecular markers and genetic diversity assessment. In addition, we describe the development of transgenic plants that have been enhanced by amaranth genes as well as new and promising gene editing techniques for amaranth, such as CRISPR and nanotechnology. Furthermore, we highlight the main germplasm and genomic databases for amaranth, as well as bioinformatic tools for assessing and analyzing the data. Altogether, this information may aid the rational design for nutritional quality and yields in *Amaranthus* and other plants, by using amaranth as a role model for nutritional improvement design.

Keywords

Amaranthus · Gene editing · Biofortification · Bioinformatics · Genetic diversity · Nutrition · Malnutrition · Food security

1 Introduction

1.1 Agricultural Importance of Amaranth

Amaranth is an ancient grain, with the first archeological register dated from 4000 BC in South America, more specifically in Mexico (Jimoh et al. 2018). Amaranth is mainly cultivated in South America (yielding 4600–7200 kg/ha), Africa (50–2500 kg/ha), and Europe (1200–6700 kg/ha) (D'Amico and Schoenlechner 2017). Amaranth

is widely grown in Africa (Ma et al. 2021), and it has been domesticated in the Americas (Stetter et al. 2016). There are three *Amaranthus* species cultivated for grain production: *Amaranthus caudatus*, *Amaranthus cruentus*, and *Amaranthus hypochondriacus* (Jimoh et al. 2018; Stetter et al. 2016).

Amaranth is a flowering plant belonging to the Amaranthaceae family and *Amaranthus* genus and includes more than 60 species (Sauer 1957; Tabio-García et al. 2021). Amaranth is considered a superfood, rich in many macro- and micronutrients necessary for humans (Jimoh et al. 2018). Besides, amaranth can grow in harsh environmental conditions, making the species very suitable for agriculture in a vast geographic range (Jimoh et al. 2018). Amaranth presents resistance to water deficit, saline soils, and acid/alkaline stresses (Tabio-García et al. 2021). Amaranth's resistance to drought and increased photosynthesis effectiveness are due to the C_4 photosynthetic pathway (Jimoh et al. 2018). C_4 plants can use CO_2 and water more efficiently, lowering photorespiration, compared to C_3 plants, like many cereals (Jimoh et al. 2018). Therefore, amaranth tolerates heat and drought, which are typical of arid and semiarid regions (Jimoh et al. 2018). Besides being resistant to abiotic stresses, amaranth is also resistant to biotic stresses (Packard et al. 2021).

Due to amaranth's resistance to several stresses, ability to thrive in a vast range of environments, and great nutritional value, this vegetable is suitable to supply high-quality foods in countries affected by hunger, such as in South Africa (Emmanuel and Babalola 2021; Ma et al. 2021). Even though amaranth has been cultivated in South Africa and many other regions, this crop has been underutilized (Emmanuel and Babalola 2021; Ma et al. 2021) due to the lack of improved, high-yielding varieties, as we will further discuss.

1.2 Amaranth Is a Rich Source of Nutrients

The three main macronutrients for the human diet are protein, carbohydrates, and fat (Holesh et al. 2022). Amaranth has high levels of the majority of the nutrients required for a balanced diet, exceeding many staple crops, such as barley, corn, rice, maize, and wheat (Adhikary et al. 2020; Joshi et al. 2018). Amaranth presents high-quality essential nutrients, such as protein, lipids, carbohydrates, vitamins, and minerals (Table 1).

Protein intake is a very important part of the human diet, providing the essential amino acids (EAA) that the human body needs, but does not synthesize itself (Floret et al. 2021; Lopez and Mohiuddin 2021). The proteins are the major effector of the cells, participating in signaling cascades, gene expression regulation, and several enzymatic reactions (Floret et al. 2021). Besides, amino acids are the building blocks for enzymes, antibodies, hormones, muscle, and many other cellular processes (Lopez and Mohiuddin 2021).

One way to determine the nutritional value of a food item is by measuring the protein content, on which if some essential amino acid is missing, it can be defined as the "limiting amino acid" (Jimoh et al. 2018). By definition, a complete protein source contains all the essential amino acids (Lopez and Mohiuddin 2021). This

Table 1 Nutrient composition from amaranth grains

Nutrient	Nutrient Type	Value per 100 g	Unit
	Water	11.29	g
	Energy	371	kcal
Protein	Total protein	13.56	g
	Glutamic acid	2.26	g
	Glycine	1.64	g
	Aspartic acid	1.26	g
	Serine	1.15	g
	Arginine	1.06	g
	Leucine	0.88	g
	Alanine	0.80	g
	Lysine	0.75	g
	Proline	0.70	g
	Valine	0.68	g
	Isoleucine	0.58	g
	Threonine	0.56	g
	Phenylalanine	0.54	g
	Histidine	0.39	g
	Tyrosine	0.33	g
	Methionine	0.23	g
	Cystine	0.19	g
	Tryptophan	0.18	g
Lipid	Total lipid	7.02	g
	Tocopherol, beta	0.96	mg
	Tocopherol, delta	0.69	mg
	Tocopherol, gamma	0.19	mg
	Phytosterol	24.00	mg
Carbohydrate	By difference	65.25	g
	Starch	57.27	g
	Fiber	6.70	g
	Sugars, total	1.69	g
Vitamins	Vitamin C (ascorbic acid)	4.20	mg
	Vitamin B5 (pantothenic acid)	1.46	mg
	Vitamin E (alpha-tocopherol)	1.19	mg
	Vitamin B3 (niacin)	0.92	mg
	Vitamin B6	0.59	mg
	Vitamin B2 (riboflavin)	0.20	mg
	Vitamin B1 (thiamin)	0.12	mg
Minerals	Phosphorus	557	mg
	Potassium	508	mg
	Magnesium	248	mg
	Calcium	159	mg
	Iron	7.61	mg
	Sodium	4.00	mg

(continued)

Table 1 (continued)

Nutrient	Nutrient Type	Value per 100 g	Unit
	Manganese	3.33	mg
	Zinc	2.87	mg
	Copper	0.53	mg
	Selenium	0.02	mg

Adapted from Jimoh et al. (2018)

notion holds particular significance when examining vegetarian diets since the most complete protein sources are usually derived from animals (Jimoh et al. 2018; Lopez and Mohiuddin 2021). On the other hand, plant-based foods are usually considered incomplete sources of proteins, for which they may lack some essential amino acid or the limiting amino acid is present in a very low concentration that it does not meet the requirements for human nutrition (Lopez and Mohiuddin 2021). Consequently, it is very important to find good sources of vegetables and cereals for a well-balanced vegetarian-based diet. Plant-based food promotes many benefits to human health, because they are a healthy source of macronutrients and dietary minerals, reducing calorie density and cholesterol consumption (Ivanova et al. 2021). The amaranth symbolizes a nearly complete source of protein for humans. For instance, amaranth grains are abundant in protein (15% of the content) and have nearly all the essential amino acids required for human nutrition (Adhikary et al. 2020)**.** Amaranth is especially rich in lysine, which is a limiting amino acid in other grains (Gebreil et al. 2020; Sisti et al. 2019). Amaranth's only limiting amino acid is leucine, which can be easily obtained by combining amaranth with other food sources (Adhikary et al. 2020). Besides its nutritional value, amaranth proteins also have antioxidant capacity (Coelho et al. 2018), and it has been considered a good source of antihypertensive peptides (Nardo et al. 2020).

On the other hand, carbohydrates have been considered a "great villain" of a healthy nutrition. However, carbohydrates are a great source of energy and, therefore, a fundamental component of the human diet (Holesh et al. 2022). Carbohydrates are categorized based on both the composition of sugars present in their structures and the specific types of sugars they contain (Holesh et al. 2022). Regarding the sugar content, they are classified into: monosaccharides (single sugar molecules), disaccharides (two sugars molecules), oligosaccharides and polysaccharides (more than two sugar molecules linked in long chains, with polysaccharides being the longest) (Holesh et al. 2022). Regarding the sugar type, they are classified as simple (one or two types of sugars), complex (three or more types of sugars), starches (complex carbohydrates rich in glucose) and fiber (non-digestible complex carbohydrates) (Holesh et al. 2022). Plants are one of the richest sources of all types of carbohydrates, with starches and fiber being very specific to these organisms (Holesh et al. 2022). A healthy and balanced diet should favor complex carbohydrates (such as whole grains) and dietary fiber, which are low in glycemic index and calories. These features are important for preventing diabetes, obesity, and inflammatory-mediated diseases (Ivanova et al. 2021). A highlight should be given

to the maintenance of the gut microbiota, which is related to many aspects of human health, such as immunity and energy metabolism (Holesh et al. 2022; Valdes et al. 2018). The consumption of low-digestible carbohydrates (LDC), present in processed foods, can be very detrimental to the gut health (Hollie et al. 2009). The LDC tends to accumulate in the large intestine, serving as substrates for pathogenic bacteria fermentation leading to the dysregulation of the gut microbiome and many other health issues (Hollie et al. 2009). Therefore, the intake of natural or unprocessed food is the best choice for a healthier gut. For instance, amaranth is a great source of healthy carbohydrates. Amaranth is rich in starch, ranging from 47% to 72% of the grains, depending on the species (Coelho et al. 2018; Nardo et al. 2020). Amaranth starch is mainly formed of amylopectin, and the unusual characteristics of the starch granules, such as small size, low resistance, and granular structure, make amaranth's grains very unique (Coelho et al. 2018). Amaranth is also rich in insoluble fiber, such as lignin and cellulose, which contributes to lowering cholesterol and to preserving gut health (Coelho et al. 2018).

Additionally, amaranth is a great source of unsaturated fatty acids, tocopherols, squalene, phytosterols, and polyphenols (Jimoh et al. 2018; Sisti et al. 2019). For instance, amaranth oil has a reasonable amount of squalene, when compared with other vegetable oils, such as olive oil (Coelho et al. 2018; Jimoh et al. 2018). Squalene is a precursor of all steroids in animals and plants (Coelho et al. 2018) and has many health properties, such as anticancer, antioxidant, and anti-inflammatory (Jimoh et al. 2018; Joshi et al. 2018). Squalene from fish oils (especially shark liver oil) has been consumed in high amounts (Kim and Karadeniz 2012). However, due to the regulations on the consumption of animal products, amaranth oil could be a great alternative source for squalene to fish oil (Ibrahim et al. 2020; Jimoh et al. 2018). Amaranth is also rich in tocopherols which are important precursor molecules for vitamin E (Ryan et al. 2007).

Aside from the macronutrients, amaranth is also rich in micronutrients, such as vitamins and minerals (Jimoh et al. 2018). Worldwide, human diets are deficient in vitamins and minerals, especially magnesium (Jimoh et al. 2018). Children and pregnant women are most affected by it, resulting in many diseases (Jimoh et al. 2018). Some of the micronutrients found in amaranth leafy vegetables and/or grain are phosphorus, iron, calcium, potassium, and vitamin A, B, C, and E (Table 1) (Jimoh et al. 2018). Therefore, amaranth is a very promising plant that can be used by humans to replenish their food needs, as well as promoting their health (Jimoh et al. 2018). Amaranth is a very promising vegetable for human consumption because of its unique nutritional and medicinal properties, as well as its ability to grow in non-abiotic and biotic soil conditions throughout the world (Gebreil et al. 2020; Jimoh et al. 2018).

1.3 Amaranth and Nutrigenomics

Food choices affect our body on many levels, including the molecular one. Nutrigenomics is the science that integrates how food constituents affect the

genomics, transcriptomics, proteomics, and metabolomics of an individual (Berná et al. 2014). Diseases occur when environmental conditions act on a predisposed genome. These external factors can be related to healthy or unhealthy lifestyles (Berná et al. 2014). This means that what we eat may interfere with gene expression and downstream signal transduction and ultimately prevent or increase the risk of diseases (Berná et al. 2014).

There is a lack of information on how amaranth-specific nutrients work at the genomic level. Besides, even though some diseases appear to have a genetic predisposition component, sometimes the pathogenesis of the disease is not very well comprehended, for instance, diabetes mellitus (DM) (Berná et al. 2014). Ultimately, it is well known that many disorders, such as obesity, cardiovascular diseases, and cancer, might be related to an interaction of predisposing genes with an imbalanced diet (Berná et al. 2014) and oxidative stress (Berná et al. 2014). Amaranth is rich in antioxidant compounds that play an important role as scavengers of the free reactive oxygen species (ROS) (Coelho et al. 2018), preventing further damage to molecules such as DNA, protein, and lipids, among others. Antioxidant molecules may as well have a role in epigenetics and nutrigenomics (Berná et al. 2014).

In order to develop personalized diets, we need to understand how epigenetics, microRNA, and others affect the expression of genes related to disease (Berná et al. 2014). A healthy diet, based on individual genomic profiles, will maximize gut microbiota health and minimize inflammation and disease susceptibility (Berná et al. 2014).

1.4 Amaranth's Importance to Prevent Diseases and Malnutrition

Plant-based food sources are a great way to combat both starvation and obesity, which pose great threats to food security.

Hunger affects the lives of billions of people worldwide, especially in Africa, and it is a great concern for all nations (Emmanuel and Babalola 2021; Webb 2021). The UN World Food Program (WFP) raised the concern that COVID-19 pandemics would accelerate and amplify the lack of food, as a result of the lockdowns to control the virus spreading (Webb 2021). The COVID-19 pandemic has increased the economic and social damage, leading to more than 70 million people to extreme poverty by the end of the year of 2020 (Webb 2021). The increased food prices amplified the consumption of lower quality food and, therefore, have increased malnutrition (Webb 2021). The COVID-19 pandemic has worsened the food crisis, but, even before that, it was estimated that almost 39% of the world population subsisted on a suboptimal diet because a healthy diet could not be afforded (Webb 2021). It is well known that simple solutions may have a great impact to improve the human diet, such as adding iodine to table salt, which prevents iodine deficiency disorders (that range from hypothyroidism to infant mortality) (Eastman and Zimmermann 2000). Likewise, introducing amaranth vegetable and/or grain as a great source of macronutrients, micronutrients, and bioactive compounds will supply requirements for a balanced diet and improve the medical conditions of consumers.

For instance, amaranth can be used for dietary fortification, improving the nutritional value of processed food (Jimoh et al. 2018). Amaranth flour could be used as a replacement for corn flour in the preparation of many bakery products, such as crackers, cookies, tortillas, and bread, increasing the nutritional value of these processed foods (Gebreil et al. 2020; Jimoh et al. 2018). Besides, amaranth flour is gluten-free, which is very important with the growth of gluten sensitivity and celiac disease (Gebreil et al. 2020; Jimoh et al. 2018). Amaranth can also be used to make pastes, sauces, dressings, and a very large variety of processed food (Jimoh et al. 2018). The high consumption of cheap processed food has been associated with overweight/obesity in both developed and developing countries (Żukiewicz-Sobczak et al. 2014). It is estimated that obesity levels and, consequently, several associated diseases are tending to increase (Żukiewicz-Sobczak et al. 2014). These diseases are more than ever in the spotlight, due to the COVID-19 pandemic (Sattar et al. 2020). People with comorbidities are more vulnerable to the acute respiratory syndrome caused by COVID-19 and also are at the greatest risk for severe illness and even death (Ivanova et al. 2021; Sattar et al. 2020). A recent study has suggested that amaranth might be important in preventing diseases, such as the COVID-19 (Datta 2021).

Amaranth compounds, from proteins to lipids and bioproducts, are potent antioxidants (Adhikary et al. 2020). Since inflammation is a major factor in many chronic diseases, a diet rich anti-inflammatory antioxidants is very important (Ivanova et al. 2021).

Amaranth compounds contribute to lowering cholesterol and consequently diminishing the chances of cardiovascular diseases (Jimoh et al. 2018; Sarker et al. 2020). For instance, amaranth fiber and protein consumption affects cholesterol metabolism and prevents hypercholesterolemia (Sisti et al. 2019). Amaranth is also a great source of vitamin E that, likewise squalene, can contribute to lower cholesterol. Vitamin E also has antioxidant, anticancer, and cardio-protectant properties, besides preventing Alzheimer's (Ryan et al. 2007). Amaranth is also rich in phytosterol, which is a plant sterol. Plant-based diets may lower cholesterol levels and reduce the risk of cardiovascular disease because they reduce sterol absorption by the digestive system (Ryan et al. 2007). Amaranth also has antihypertensive peptides (Nardo et al. 2020).

Amaranth is rich in unsaturated fatty acids, such as omega-3 and omega-6 (Soriano-García et al. 2018). Amaranth lipid portion, such as squalene, has antioxidant, autoinflammatory, and anticancer properties (Soriano-García et al. 2018). Besides the aforementioned medical properties, amaranth natural antioxidants may also have a role in the prevention of cataracts and retinopathies, arthritis, atherosclerosis, and even neurodegenerative diseases (Sarker et al. 2020). The bioactive compounds of the lipid fraction may additionally have antidiarrheal and antidepressant effects as well (Coelho et al. 2018).

Great attention has been given to the *Amaranthus* component betalain, which is responsible for amaranth red color (betacyanin) (Adhikary et al. 2019) (Fig. 1) and can be used as a natural colorant for food (Coelho et al. 2018). Betalains have antioxidant, anticancer, antibacterial, and antimalarial properties (Packard et al. 2021). The betalain from amaranth is therefore a potential natural coloring agent with interesting health benefits for the food industry (Tabio-García et al. 2021). Consumers who want a

Fig. 1 *A. tricolor* "Red Leaf"

healthier lifestyle and are concerned about synthetic colorants' detrimental effects have been attracted to these features (Tabio-García et al. 2021).

There are many nutrients in amaranth, but some compounds, such as non-digestible glucosides, can reduce its ability to be digested and utilized (Jimoh et al. 2018). However, it is possible to improve the bioavailability of nutrients in the edible parts of plants through biofortification (Malik and Maqbool 2020). In the next section, we will discuss bioavailability and biofortification in more detail.

1.5 Bioavailability of Amaranth Nutrients

Bioavailability is the nutrient portion that is utilized by the human body (Jimoh et al. 2018). The fact that amaranth contains a diversity of compounds does not mean all of them are absorbed by the body. For instance, the bioavailability of amaranth proteins can be reduced by the presence of antinutrient compounds, such as phytic acid, nitrate, oxalates, and tannins (Coelho et al. 2018; Jimoh et al. 2018). So far, these compounds have not been associated with toxic effects on human nutrition (Coelho et al. 2018). Therefore, amaranth grain should be properly processed before

ingestion, to increase bioavailability (Coelho et al. 2018; Jimoh et al. 2018). Another trending field is using encapsulation to improve bioavailability of amaranth bioactive compounds (Coelho et al. 2018).

1.6 Amaranth Biofortification

Micronutrients, such as iron, zinc, iodine, and vitamins, are crucial for human health, and their lack in malnourished populations can lead to many disorders and diseases (Malik and Maqbool 2020). In countries with a limited choice of diets, biofortification holds the promise of providing a single source of complete nutrition (Malik and Maqbool 2020). Some health conditions affected by malnourishment are highlighted in Table 2. Pregnant women and children are particularly vulnerable to these conditions (Malik and Maqbool 2020).

Biofortification can be obtained through agricultural practices, traditional breeding, or genetic editing (Malik and Maqbool 2020). Staple cereals, vegetables, beans, and fruits have been biofortified using these approaches (Garg et al. 2018). Different from simply adding nutritional value to staple foods, biofortification attempts alterations in the plant phenotype (Malik and Maqbool 2020). Besides creating grains with enhanced nutritional value, the biofortified plants also may become less susceptible to biotic and abiotic stresses (Malik and Maqbool 2020).

So far, our agricultural practices have not been focused on increasing nutrient quality but, instead, maximizing yield, which can decrease the nutritional quality of the grains (Garg et al. 2018). Agricultural practices that can favor crops' biofortification involve fertilization, seed treatment and nurseries, and certified seeds (Ochieng et al. 2019).

1.6.1 Genetic Approaches for Amaranth Biofortification

There is a great effort from the scientific community to produce plants with an improved protein profile, including limiting amino acids (Wenefrida et al. 2013).

Table 2 Micronutrient deficiency and health conditions

Deficiency in micronutrients	Health condition
Iron	Anemia
Zinc	Night blindness, dermatitis, loss of appetite, impaired immune system
Iodine	Pregnancy related issues (affects the neurodevelopment of fetus), affects growth and cognitive functions in children
Vitamin A	Blindness, pregnancy and breastfeeding related issues, impaired immune system
Vitamin B	Skin inflammation, impaired immune system, fatigue, and depression
Vitamin C	Dermatologic issues, joint pain, and impaired immune system
Vitamin E	Nerve and muscle damages, impaired immune system, hemolytic anemia, ophthalmological disorders

Source: Malik and Maqbool (2020)

Several methods can be used with this purpose, such as genetic breeding and genetic engineering, genomics, and selection assisted by markers (Wenefrida et al. 2013).

Biofortification through genetic breeding is possible when genetic diversity is available for a given species (Garg et al. 2018). Genetic diversity is represented in the germplasm databases, which are a collection of genetic resources for a given organism. Amaranth includes up to more than 11,000 accessions among 8 gene banks (Joshi et al. 2018). However, it seems that amaranth genetic diversity has been poorly used in breeding programs (Joshi et al. 2018).

Finally, genomic editing is one of the most promising approaches to improving amaranth nutritional value. The increased amount of genetic data generated by sequencing, such as amaranth reference genome (Lightfoot et al. 2017) and other genomic and transcriptomic data (Clouse et al. 2016; Délano-Frier et al. 2011; Sunil et al. 2014), is increasing the potential to edit amaranth genome.

Using gene edition, amaranth macronutrients, such as protein, have been targeted (Garg et al. 2018). The amaranth lysine pathway, for example, is an attractive biofortification target. According to Sunil et al. (2014), amaranth contains more lysine and threonine than major cereals such as rice and wheat, and some plants don't even produce lysine or threonine (Wenefrida et al. 2013).

Using amaranth genes, many crops have been biofortified. To modify rice grains, which lack lysine and threonine, synthetic genes combining endogenous rice sequences with lysine and threonine coding sequences were inserted under the promoter 35S to produce these amino acids in large quantities (Jiang et al. 2016). However, there are limitations to the production of transgenic biofortified plants (Jiang et al. 2016). Still, Jiang et al. (2016) research opened venues for producing transgenic plants with synthetic genes with enhanced lysine and threonine content. In order to produce plants biofortified with lysine and threonine, we must understand their biosynthetic pathway. In this regard, the draft of *Amaranthus hypochondriacus* genome and transcriptome was obtained in 2014 (Sunil et al. 2014), with particular attention given to the annotation of the genes involved in lysine biosynthesis. Amaranth genes such as albumin genes (AmA1) have also been used to change the protein content in plants, such as wheat (*Triticum aestivum* L.) (Garg et al. 2018). The amino acids that increased in content were methionine, cysteine, and tyrosine, besides lysine (Garg et al. 2018; Tamás et al. 2009). In addition, transgenic potatoes were also improved in amino acid content using the same albumin gene (AmA1) from amaranth (Chakraborty et al. 2010; Garg et al. 2018). A further result of this biofortification was an increase in potato photosynthetic activity and yield, proving the importance of biofortification with the AmA1 albumin gene from amaranth (Chakraborty et al. 2010).

It appears that amaranth genes have been used more to biofortify plants than to alter the amaranth itself. It is likely that the most recent gene editing strategies will increase the possibility of improving amaranth traits in the near future.

1.7 Biochemical Pathways of Amaranth Compounds

Understanding the genes and proteins underlying a biosynthetic pathway is crucial to improving plant breeding programs using gene editing.

This chapter will focus on two amaranth compound biosynthetic pathways, which may be of great interest for the food and health industries: lysine biosynthesis and betalain biosynthesis.

The lysine biosynthetic pathway in plants shares a common route with the amino acids threonine, isoleucine, and methionine: the aspartate pathway, where L-aspartate is the precursor of these amino acids (Sunil et al. 2014). Two main enzymes regulate the lysine biosynthetic pathway: the aspartate kinase (AK; EC 2.7.2.4) and the dihydrodipicolinate synthase (DHDPS; EC 4.3.3.7) (Sunil et al. 2014). As stated by Sunil et al. (2014), the high-lysine content in *Amaranthus* is due to the presence of only one ortholog of aspartate kinase 1 (AK1) gene and high expression of the dihydrodipicolinate synthase (DHDPS) gene in seeds.

The lysine biosynthetic pathway in amaranth is not very clear yet. Sunil et al. (2014) published a draft of *A. hypochondriacus* genome and found several isoenzymes AK and DHDPS, among other five important enzymes of the aspartate biosynthetic pathway in plants, such as aspartate semialdehyde dehydrogenase (ASD; EC 1.2.1.11), dihydrodipicolinate reductase (DHDPR; EC 1.17.1.8), diaminopimelate aminotransferase (DAPAT; EC 2.6.1.83), diaminopimelate epimerase (DAPE; EC 5.1.1.7), and diaminopimelate decarboxylase (DAPDC; EC 4.1.1.20). The gene expression profile of the AK and DHDPS revealed a correlation with the high-lysine content in amaranth (Sunil et al. 2014). The study of amaranth lysine biosynthesis as a model for understanding plant lysine biosynthesis is of great importance in order to produce high-lysine content plants.

Regarding betalains, these molecules are composed of at least one heterocyclic ring containing a nitrogen atom and are derived from betalamic acid [4-(2-oxoethylidene)-1,2,3,4- tetrahydropyridine-2, 6-dicarboxylic acid] (Böhm and Rink 1988; Packard et al. 2021). The betalain biosynthetic pathway is controlled by several factors and conditions, and it was very well discussed by Packard et al. (2021). In summary, betalamic acid condensates with *cyclo*-DOPA, forms betacyanin, and, when it condensates with amino acids or amines, forms betaxanthin (Böhm and Rink 1988; Packard et al. 2021). L-tyrosine is the main precursor of betalains, being first converted to L-DOPA and subsequently to cyclo-DOPA by several enzymes of this biosynthetic pathway (Böhm and Rink 1988; Packard et al. 2021).

The betalain biosynthetic pathway has been carefully reviewed by Packard et al. (2021), and it is a reliable source of information to guide gene editing or gene expression of betalain-related genes. It is interesting to notice that betalains are not very well represented among many plant edible species; therefore, they are less consumed than other antioxidant pigments (Grützner et al. 2021). Grützner et al. (2021) suggested that incorporating betalain biosynthetic genes in a varied range of transgenic species, such as tomatoes, could increase betalain consumption. Moreover, many other transgenic plants producing betalain have been obtained through heterologous expression (Packard et al. 2021). Therefore, these techniques can be exploited to scale up betalain production commercially (Packard et al. 2021).

Since many plants lack the amino acid lysine, and the betalain pigment is restricted to the order of Caryophyllales, amaranth is an excellent role model for

understanding the biosynthesis of these two compounds, which present many desirable effects for human nutrition and health.

2 Genetic Resources of Amaranth

2.1 Brief on Genetic Diversity Analysis of Amaranth

Amaranth and its wild relatives harbor considerable phenotypic plasticity and genetic variability (Gerrano et al. 2017) which contribute to their adaptability to a variety of soil and agroclimatic conditions (Dharajiya et al. 2021). Therefore, amaranth has excellent potential for breeding and variety development. In this respect, the characterization of genetic diversity of amaranths is fundamental for the development of germplasm widely diverse to be efficiently used in genetic breeding programs (Dharajiya et al. 2021; Thapa et al. 2021). As previously described, amaranth grain has high potential from cultivation in the field to the production of value-added products, including nutritional and health benefits (D'Amico and Schoenlechner 2017).

The amaranth genus has gained attention as a promising food crop, mainly owing to its tolerance to stressful conditions (Pulvento et al. 2021) besides its great nutritional and functional benefits (Coelho et al. 2018; D'Amico and Schoenlechner 2017). Moreover, in view of the wide geographic distribution and genetic diversity, cultivated and weedy amaranths emerged as appealing models to be explored in improving yield and nutraceutical value (Soriano-García et al. 2018).

Efforts have been made to maintain amaranth's diversity in germplasm banks worldwide (Das 2016; Joshi et al. 2018). Examples of germplasm banks include the US National Plant Germplasm System, with more than 3000 accessions of amaranth from 40 countries (Joshi et al. 2021; Trucco and Tranel 2011). Various morphological and molecular markers have been used for the study of genetic diversity and evolutionary relationships in selected species of the genus *Amaranthus*. We will further discuss the use of molecular markers in the characterization of the amaranth germplasm.

2.1.1 Genetic Diversity Analysis of Amaranth Assisted by Morphological Traits

An approach adequately used to assess genetic diversity is through a combination of morphological, biochemical, and molecular data within amaranth genus (Govindaraj et al. 2015). Morphological descriptors (i.e., color of leaf, flower, stem, and panicle; leaf and panicle shape; plant habit; etc.) have been used to assess amaranth's diversity (Gerrano et al. 2017; Thapa and Blair 2018). Such characteristics are useful to obtain basic information on existing morphological variability from cultivated species and their wild relatives, which can be incorporated into pre-breeding programs (Showemimo et al. 2021). Examples include the selection of parental materials based on the existence of genetic variation. Gerrano et al. (2017) used morphological traits of *Amaranthus* species, aiming to identify the best parents for

the traits of interest that could be used in the *Amaranthus* breeding program. This study revealed a wide genetic diversity among the species evaluated, helping with the selection of target accessions for the amaranth breeding program in South Africa (Gerrano et al. 2017). Thapa and Blair (2018) observed a wide diversity in phenotypic traits among amaranth species when studying 293 cultivated grain amaranths and their wild relatives. According to the authors, field assessment of major morphological traits can be successful in amaranth grain with implications for breeding varieties (Thapa and Blair 2018). In another study (Showemimo et al. 2021), 12 amaranth grain accessions exhibited high phenotypic and genetic variability for all the traits measured (particularly the desired leaf traits and grain yield traits). Phenotypic variances were slightly higher than the genotypic variance for all the traits measured, revealing little influence of the environment, suggesting that the accession variability is considered useful for amaranth selection and improvement (Showemimo et al. 2021).

Morphological characters also have been useful in understanding the evolutionary relationship of *Amaranthus* species. In India, a study revealed an evolutionary relationship of six amaranth species (*A. blitum*, *A. blitoides*, *A. deflexus*, *A. dubius*, *A. polygonoides*, and *A. retroflexus*) from different states (Kamble and Gaikwad 2021). However, the insufficient distinctive morphological characters for *Amaranthus* species have hampered the use of morphological markers alone (Govindaraj et al. 2015; Thapa and Blair 2018). For this purpose, molecular markers have been widely considered to assess *Amaranthus* genetic diversity in complementing morphological information (Oduwaye et al. 2019).

2.1.2 Genetic Diversity Analysis of Amaranth Assisted by Molecular Markers

Germplasm characterization based on molecular markers has gained importance due to the speed and quality of information of genetic diversity data created. The use of improved marker systems may aid in the better characterization of the vegetal germplasm, rather than using only the morphological trait system (Dar et al. 2019). For this purpose, molecular markers have been employed in germplasm characterization, phylogenetic analysis, and genetic diversity of the major crop species (Dar et al. 2019). Markers such as amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), microsatellites or simple sequence repeats (SSRs), inter-simple sequence repeat (ISSR), and single nucleotide polymorphisms (SNPs) have been efficiently used to evaluate the variability of germplasm accessions in amaranth (Thapa and Blair 2018).

For instance, AFLP was applied by Chandi and collaborators (2013), to assess genetic diversity in Palmer amaranth populations (from North Carolina and Georgia) and to characterize diversity among states, populations, gender, and even response to glyphosate. These results are important, for example, to optimize weed management in breeding programs of amaranth populations (Chandi et al. 2013).

The RAPD marker has also been used to assess genetic diversity and in the selection of potential parents for characters of interest in plant breeding programs. For instance, the genetic diversity within 29-grain amaranth accessions in Southwest

Nigeria was accessed using 27 phenotypic characters (10 morphological and 17 nutritional) and RAPD marker (Akin-Idowu et al. 2016). The level of polymorphism obtained with RAPD primers (81%) broadly indicates the degree of heterogeneity observed in this amaranth germplasm (Akin-Idowu et al. 2016). The RAPD marker has also contributed to the understanding of the origin and evolution of cultivated amaranth and wild species. In this sense, seven amaranth species from different phytogeographic regions were evaluated by RAPD which was sufficient to promote a level of (intra- and interspecific) informative characters (Kumar Pandey et al. 2019).

Furthermore, the availability of high-throughput sequencing technologies has facilitated the fast discovery of thousands of microsatellite regions, which may further facilitate understanding evolutionary relationships, genetic diversity, and genomic texture of *Amaranthus* species (Tiwari et al. 2021).

Due to their multiallelic nature, codominant inheritance, and wide distribution throughout the plant's and animals' genome, including organellar genomes (mtSSR and cpSSR) (Gupta et al. 2021), microsatellites have gained considerable importance in plant genetic diversity studies, population genetics, and evolutionary studies. They have also been used in fundamental research such as genome analysis, gene mapping, and marker-assisted selection, among others (Gupta et al. 2021).

Using *A. hypochondriacus* genome, Tiwari et al. (2021) identified and characterized a set of SSRs, which facilitated the identification of 97 alleles among 10 *Amaranthus* species (*A. hypochondriacus*, *A. caudatus*, *A. retroflexus*, *A. cruentus*, *A. tricolor*, *A. lividus*, *A. hybridus*, *A. viridis*, *A. edulis*, and *A. dubius*), revealing relationships among amaranth spp. that could be useful in species identification, DNA fingerprinting, and QTL/gene identification (Tiwari et al. 2021).

Amaranthus genetic diversity has also been accessed by chloroplast sequences**.** Full-length chloroplast sequences for four *Amaranthus* species (*A. hypochondriacus*, *A. cruentus*, *A. caudatus*, and *A. hybridus*) were assembled by Chaney et al. (2016) revealing polymorphic microsatellite and informative SNP (*single nucleotide polymorphism*) markers, with application in amaranth phylogenetic and genetic diversity studies. A chloroplast matK marker (maturase K gene) was used to identify leafy amaranth species from Vietnam (Nguyen et al. 2019). Twenty-one SSR markers were developed for *A. tricolor* 'Biam', which successfully amplified single alleles in 294 Vietnamese and overseas accessions (Nguyen et al. 2019). According to the authors, there was a positive relationship between geographic distance and genetic differentiation among most of the overseas groups and the Vietnamese collection (Nguyen et al. 2019). In another study, Hong et al. (2019) also related a close relationship among the chloroplast genomes of *Amaranthus hybridus*, *A. hypochondriacus*, and *A. caudatus*. This relationship was reassured by the complete chloroplast genome of *A. hybridus* sequencing (Bai et al. 2021), which revealed a close relationship between *A. hypochondriacus* and *A. caudatus*. These studies are fundamental to promoting a better understanding of the phylogeny and evolution of the genus *Amaranthus*.

2.1.3 The Evolutionary Relationship of Amaranth

Molecular markers have also made great contributions to the elucidation of the origin and evolution of amaranth. The evolutionary relationship among *A. caudatus*,

A. hybridus, and *A. quitensis* was determined by Stetter et al. (2016). In this study, 119 accessions of the 3 species from the Andean region were genotyped using genotyping by sequencing (GBS) with phenotypic variation in 2 domestication-related traits: seed size and seed color (Stetter et al. 2016). As a result, 9485 SNPs were obtained, and a strong genetic differentiation of the cultivated *A. caudatus* from the relatives *A. hybridus* and *A. quitensis* was observed (Stetter et al. 2016). Given its utility in detecting large numbers of SNPs, the GBS technology was applied to study the diversity and population structure of the grain amaranth, compared to an outgroup of *Amaranthus* species (Wu and Blair 2017). This analysis revealed a high variation within amaranth grain species and their close relatives (Wu and Blair 2017).

Thapa et al. (2021) studied the diversity and relationship among collections of amaranth (accessions of all three cultivated grain amaranth species and their wild relatives). The authors suggested that two Mesoamerican species (*A. cruentus* and *A. hypochondriacus*) were intercrossed and distantly related to the South American species (*A. caudatus* and the weedy relative *A. quitensis*), which persisted in a wild-cultivated hybrid state (Thapa et al. 2021).

The knowledge and characterization of genetic diversity, within amaranth's populations, as well as germplasm conservation, are essential steps for developing new varieties/cultivars with desired agronomic traits, and it is useful for exploring its nutritional and medicinal properties. Collectively, molecular markers have been an excellent approach for various purposes, including genetic studies, breeding efforts, and marker-assisted selection of amaranth.

3 Molecular Mapping of Genes/QTLs Underlying Nutritional Traits in Amaranth

DNA markers have been a useful tool not only for assessing intraspecific variation but also for the subsequent characterization and identification of quantitative trait loci (QTLs) for agronomic and nutritional quality in amaranth genotypes (Akin-Idowu et al. 2016). Marker-assisted breeding employs the linkage between known QTLs and genetic markers to select individuals. SNP markers have therefore been effective when used in marker-assisted selection (MAS), especially when linked to target QTLs (Stetter et al. 2016).

When it comes to complex traits, effective breeding methods are needed to identify and exploit previously unexplained trait variation caused by a large number of small-effect QTLs, as well as methods that utilize genomic selection (GS) information (Varshney et al. 2021). As high-throughput sequencing technologies have improved, crop breeding programs have become more precise and efficient due to the discovery and mapping of genome-wide allelic variation (Varshney et al. 2021). For instance, genome-wide surveys on comprehensive diversity panels and reference genomes have facilitated phenotypic associations. The DArTSeq technology (based on GBS methods) was used by Jamalluddin et al. (2022) to determine the genetic relationships and population structure between 188 amaranth accessions from 18 agronomically important vegetable, grain, and weedy species. Out of the

74,303 SNP markers identified, 63,821 could be physically mapped to the *A. hypochondriacus* genome (Jamalluddin et al. 2022). This marker allowed a genome-wide association study (GWAS) analysis for ten morphological traits and, hence, provided useful information on genetic diversity and its correlation with agronomic traits in amaranth (Jamalluddin et al. 2022).

In addition, a universal SNP dataset across *Amaranthus* species was produced using double-digest restriction site-associated DNA (ddRAD-Seq) sequencing, allowing access to genetic diversity across the major species of the World Vegetable Center (with over 1000 *Amaranthus* accessions) (Lin et al. 2022). Candidate loci that regulate days to flowering were revealed by an interspecific genome-wide association study (GWAS) (Lin et al. 2022). This study showed that genotypic data is useful for species demarcation in the genus *Amaranthus* and that interspecific GWAS can detect quantitative trait loci (QTLs) across species (Lin et al. 2022). Recently, *A. tricolor* accessions (preserved by the World Veg and USDA gene banks) were also evaluated using genome-wide SNPs developed by ddRAD-Seq (Hoshikawa et al. 2023). The 440 accessions contained 10,509 SNPs, with 5638 without missing data (Hoshikawa et al. 2023). With these results, a core collection is now available that represents a wide diversity of amaranth germplasm. This collection will help researchers find important agronomic gene loci to improve genetic breeding, thus improving the economics of amaranth (Hoshikawa et al. 2023).

In addition, Ma et al. (2021) assembled the genome of *A. cruentus* at the chromosome level using short-read, long-read, and phased sequencing technologies. The goal of this study was to identify the genomic features related to the high levels of micronutrients and proteins in these species leaves (Ma et al. 2021). In this study, gene colocations for the key enzymes that synthesize betalain were found across the Amaranthaceae family (Ma et al. 2021). Furthermore, 22 possible biosynthetic gene clusters were identified in *A. cruentus* genome (10 of which are conserved in *A. hypochondriacus*) that may facilitate the elucidation of other metabolic biosynthetic pathways present in this plant species (Ma et al. 2021). This will benefit amaranth breeding programs.

As in cultivated species, the genome sequencing of wild and weedy species would be advantageous for various large-scale genotyping applications, including germplasm characterization, cultivar identification, and QTL discovery (Joshi et al. 2018). In contrast to major crops (including model plants), amaranth weeds have not been adequately explored, and genomic tools can be efficiently applied in the identification and characterization of genes related to genetic traits that make weeds so resilient when compared with crops in identical niches. Thus, weed species of agronomic importance, such as *A. rudis* (common waterhemp), *A. palmeri* (palmer pigweed), and *A. hybridus* (smooth pigweed), are also potential candidates for genomics studies (Trucco and Tranel 2011).

The QTL regions for different traits of amaranth have been identified (Lightfoot et al. 2017; Maughan et al. 2011). Recently, Stetter and collaborators (2020) showed strong evidence that the grain amaranth was domesticated three times and that the white seed color was independently selected in each grain species. Furthermore, the authors identified a MYB-like transcription factor (TF) gene within a QTL region

showing to be a potential regulator of seed coat color variation of amaranth grain (Stetter et al. 2020).

TFs have also been related in different amaranth studies as the most promising gene family involved in response to biotic and abiotic stresses (Palmeros-Suárez et al. 2021). Therefore, TFs are important candidates to increasing plant resistance to stresses through genetic engineering (Palmeros-Suárez et al. 2021).

Genetic mapping is an important resource to gain further insight into the underlying genes for traits of economic importance. Thus, the knowledge of genetic diversity in amaranth breeding programs is essential, especially in the identification of the best parents to a generation of segregating population with genetic variability for introgression of desired traits in the development of high-yielding varieties of amaranth (Dharajiya et al. 2021). The use of genomic prediction could largely accelerate the genetic gain for nutritional traits and holds great potential for biofortification breeding in grain amaranth (Joshi et al. 2018).

4 Limitations of Conventional Breeding and Genetic Resources of Amaranth

A little number of amaranth cultivars with enhanced agricultural traits have been released so far (Schafleitner et al. 2022). The most successful amaranth variety implemented in the USA has been the Plainsman (Baltensperger et al. 1992), which is an improvement of *A. hypochondriacus* genotype, where the trait earliness was introduced (Joshi et al. 2018). However, there is still so much room for improvement of amaranth cultivars, especially if the "wild" genes from native species, which represents a diverse genetic resource, were more explored (Joshi et al. 2018).

One of the challenges to the amaranth breeding program is the nature of amaranth flowers, which are numerous and self-pollinating, making the crossing processes difficult (Stetter et al. 2016). Stetter et al. (2016) evaluated different crossing methods within three amaranth grain species, where the highest success rate was obtained with the hand emasculation method (74%). It has been suggested that amaranth heterotic breeding across the grain cultivars could also be a good strategy for advanced breeding in amaranth (Joshi et al. 2018). Additionally, mutation breeding has shown great results in increasing the polygenic variability of four cultivars of amaranth (Joshi et al. 2018). Lysine content, among other traits, has been increased in amaranth species through mutational approaches (Joshi et al. 2018).

Other traits that have been targeted for amaranth genetic breeding are plant height reduction, increased seed size, and adaptation to the mechanical harvesting (Joshi et al. 2018; Schafleitner et al. 2022; Stetter et al. 2016). Besides, improving the taste and nutritional quality of the leaves, rapid growth, and resistance to biotic and abiotic stresses are important targets to plant breeding programs (Schafleitner et al. 2022). Even though amaranth is very resistant to abiotic and biotic stresses, this trait should be kept in breeding pools to improve amaranth's abundance and distribution (Joshi et al. 2018). The knowledge of the genetic diversity of

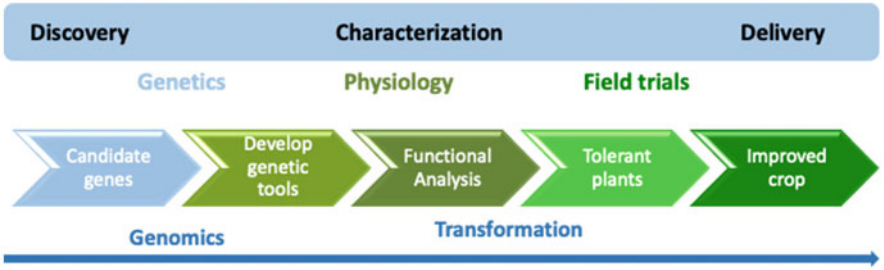


Fig. 2 Improved crop design flowchart

Amaranthus species will provide information for creating superior cultivars in plant breeding programs (Joshi et al. 2018).

Amaranth genetic traits and the current state of genetic breeding have been carefully reviewed by Joshi et al. (2018). Genetic resources, genotyping, and genomics are the core information to create superior genotypes through traditional breeding programs, genetic engineering, and genome editing (Fig. 2) (Joshi et al. 2018). This will lead to an increased use of amaranth as a main agricultural crop. The use of minor crops will increase agrobiodiversity and, therefore, favor food global security (Joshi et al. 2018).

Although genomic prediction can accelerate the selection of grain with increased nutritional value, very few studies have been performed with *Amaranthus* so far (Joshi et al. 2018) as previously discussed. The available strategies for amaranth gene editing will be discussed in the next section.

5 Strategies for Amaranth Gene Editing

Gene editing is a great strategy for precise genome modifications in plants (Wada et al. 2020). The most advanced technologies for gene editing are zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 9 (CRISPR/Cas9) (Liang et al. 2014; Wada et al. 2020). CRISPR has many advantages when compared with ZFNs and TALENs, because it is simpler, while these other techniques require more complex constructs (Bortesi and Fischer 2015; Joshi et al. 2018; Liang et al. 2014). CRISPR allows using multiple targets for genetic improvement and therefore can aid in the creation of new crop varieties in an unprecedented way, by inducing double-strand breaks (DSBs) in DNA (Wolter et al. 2019). CRISPR is a powerful tool for genome editing and is generating exciting new possibilities to increase the genetic diversity of amaranth (Joshi et al. 2018). However, none of these genome editing techniques have been used with amaranth so far (Joshi et al. 2018), even though many other studies using reverse genetics have been published for this plant (Packard et al. 2021).

Amaranth is a great candidate for gene editing since it has a vast repository of genomic data (Clouse et al. 2016; Délano-Frier et al. 2011; Lightfoot et al. 2017; Sunil et al. 2014), and protocols for regeneration using tissue culture are well-established for this plant (Adhikary et al. 2019; Joshi et al. 2018). In fact, the technology VIGS (virus-induced gene silencing) has been successfully used to study genes of betalain biosynthesis in amaranth (Adhikary et al. 2019). Adhikary et al. (2019) work established an efficient genetic transformation protocol that can be used for the improvement of genes involved with amaranth's nutritional and medicinal properties, besides the yield (Joshi et al. 2018). That being said, it is very important to invest in the development of protocols for using the most varied gene editing techniques in amaranth.

Another potential technology for amaranth genome editing is the nanotechnology-based genome editing system (Ahmar et al. 2021). This technology claims to be even more efficient, safe, and simpler than CRISPR (Ahmar et al. 2021). It is also possible that CRISPR/Cas9 nanotechnology coupled systems represent the most advanced technology available in gene editing (Naik et al. 2022). These techniques have not been employed in amaranth so far but could speed up the creation of improved varieties at a lower cost, by overcoming the challenges of traditional breeding methods (Ahmar et al. 2021). In conclusion, the production of varieties with high nutritional value and yields, and many times also resistant to biotic and/or abiotic stresses, would be an incredible mark for food safety and in the combat of malnourishment worldwide.

6 Brief Account on the Role of Bioinformatics as a Tool

For some years, biology and computer science were very connected. Therefore, a new field has emerged named bioinformatics, which is the science of managing, mining, integrating, and interpreting information from biological data from different levels of an organism.

Given the complexity and exponential growth of biological data from omics projects, computational storage and analysis techniques have become increasingly demanded, to solve three main problems: efficient storage, information's management, and extraction of relevant information from the data (Quintans et al. 2022). The last problem is one of the main challenges of computational biology and bioinformatics, which requires the development of tools/databases capable of transforming all this heterogeneous data into biological knowledge (Greene et al. 2014). These bioinformatic databases and tools go beyond a description of the data, by delivering new biological insights. It is almost impossible to carry out research in bioinformatics without using a database and bioinformatic tools to process and explore the vast collection of available data, both public and private.

In this context, in this sections, we will describe databases of genes, genomes, gene expression, proteins, genome comparison tools, and the integration of other different types of data from plants of the amaranth genus that allow increasing knowledge about the system biology of these plants, providing nutritive and weed

resistance biotechnological advances. Furthermore, the amaranth genome study will underpin the research of this orphan crop and accelerate the development of improved varieties.

6.1 Amaranth Genome Sequencing

To date, a few amaranth genomes have been published, such as *A. hypochondriacus* (Clouse et al. 2016; Sunil et al. 2014) and *A. cruentus* (Clouse et al. 2016; Ma et al. 2021) of grain species and *A. tuberculatus*, *A. hybridus*, and *A. palmeri* (Montgomery et al. 2020) of weed species. Table 3 summarizes the genomic information of amaranth, such as haploid chromosome number, total genome length (Mb), and the GC content.

A. hypochondriacus genome was sequenced in 2014 (Sunil et al. 2014) representing the first C4 dicot genome and the first grain species from Caryophyllales. *A. hypochondriacus* is an interesting species for its C4 metabolism, being capable of tolerating many adverse conditions, such as drought, and still producing highly nutritious seeds (NCBI Resource Coordinators 2016; Sunil et al. 2014). *A. hypochondriacus* genome encodes at least 24,829 proteins (Sunil et al. 2014). However, only 112 gene results are displayed after a search at NCBI, when the query "*Amaranthus hypochondriacus*" is used in the "Gene" category (NCBI Resource Coordinators 2016, Access 24 June, 2022). Likewise, a high-quality draft of *A. hypochondriacus L.* genome (N_{50} of 371 kb) was obtained by Clouse et al. (2016) using whole-genome sequencing (WGS) and transcriptome sequencing (TS). The assembled genome presented 377 Mb, being rich in repetitive sequences (48% of the sequences), including many retrotransposons (Clouse et al. 2016). In this study 23,059 protein-coding genes were identified (Clouse et al. 2016) corroborating with Sunil et al. (2014). Clouse et al. (2016) also acknowledged that *A. hybridus* is the progenitor species of the grain amaranth.

Table 3 Summary of information about the genomes of *Amaranthus* species: haploid chromosome, total length (Mb) of genome, and GC content

	Amaranthus hypochondriacus (Sunil et al. 2014)	*Amaranthus hypochondriacus* (Clouse et al. 2016)	*Amaranthus tuberculatus* (submitted by the University of Illinois)	*Amaranthus palmeri* (submitted by Clemson University)	*Amaranthus cruentus* (Ma et al. 2021)
Haploid chromosomes	16	16	16	17	17
Total length (Mb)	417.46	377	572.9	395.293	370.914
GC%	34.5087	34.7	35	33.45	33.0875

Source: (CABI 2022; Clouse et al. 2016; Ma et al. 2021; Montgomery et al. 2020; Sunil et al. 2014)

Furthermore, a genome draft of three *Amaranthus* weed species of agricultural importance was published in 2020: *A. tuberculatus*, *A. hybridus*, and *A. palmeri* (Montgomery et al. 2020). Montegomery and collaborators (2020) obtained the most contiguous draft assemblies for these amaranth weeds, including the very first genome assembled for *A. hybridus*, by combining several sequencing approaches, such as long-read sequencing and chromatin mapping (Montgomery et al. 2020). The study of the weed's genome will aid in a better understanding of herbicide's resistance evolution, which is essential for weed's management (Montgomery et al. 2020).

In 2021, another amaranth grain species, *A. cruentus,* had its genome sequenced (Ma et al. 2021). *A. cruentus* genome (370.914 Mb and 33.0875% of GC content) was sequenced using short-read, long-read, and phased sequencing technologies by the University of York (Ma et al. 2021). However, *A. cruentus* lacks details regarding the genome annotation, and also, there is no gene associated with this species in NCBI (Ma et al. 2021; NCBI Resource Coordinators 2016).

Currently, the NCBI holds 1109 plant genomes and 476,498 plant genes, of which 518 are from *Amaranthus* species (NCBI Resource Coordinators 2016, Access 31 March, 2022). There are four species of *Amaranthus* with their whole genomes sequenced and available at the NCBI, namely, *A. hypochondriacus*, *A. tuberculatus*, *A. palmeri*, and *A. cruentus* (NCBI Resource Coordinators 2016, Access 31 March, 2022).

The amaranth genome sequencing supplies a reference genome for transcriptomic and proteomic studies, besides being valuable for evolutionary insights of *Amaranthus* genus, the Caryophyllales order, and, ultimately, angiosperms in general (Sunil et al. 2014; Adhikari et al. 2021). This knowledge will be useful for the rational design of amaranth crops, with increased productivity and quality traits (Ma et al. 2021).

6.2 Amaranth Sequences Among the Gene and Genome Databases

Genomic databases are DNA sequence repositories from many different species of plants and animals. The National Center for Biotechnology Information (NCBI), for instance, is one of the most important biological databases and stores genes from different organisms in domains of life (NCBI Resource Coordinators 2016).

There are many types of databases according to the biological data, such as genomic databases (DNA sequences), transcriptomic databases (RNA sequences), proteomic databases (protein sequences), and structure databases (mainly protein structures). There are also more specific databases, such as Metabolic and Signaling Pathways, Diseases Databases, Gene Expression Databases, Plant Databases, and Taxonomic Databases. Each of them includes a specificity, and some can integrate a variety of database functions (Quintans et al. 2022). These databases offer information on the sequences and can be created for specific organisms, such as animals and plants. For instance, the AmaranthGDB (Gonçalves-Dias and Stetter 2021) is a genetics and genomic database specific to the genus *Amaranthus*.

6.2.1 AmaranthGDB

AmaranthGDB is a database with genetic and genomic information on amaranth and tools to accelerate amaranth research (Gonçalves-Dias and Stetter 2021). One of the available tools is the PopAmaranth, which is a genetic browser that combines amaranth genomic data and genome-wide population (Gonçalves-Dias and Stetter 2021). The information is easily retrieved and may be used to identify target genes by the researcher (Gonçalves-Dias and Stetter 2021). Regarding improving the nutritional value of amaranth, some genetic features have been pinpointed as putative targets, such as genes that participate in the synthesis of phytic acid (which is an antinutrient), gene families of ion transporters, and biosynthetic gene clusters within amaranth line (Gonçalves-Dias and Stetter 2021).

6.2.2 Amaranth and Gene Expression Databases

The genomic projects have been adding expressive data to the current databases, as we previously discussed. Although genomic data are important, the identification of genes linked to important agronomic traits requires gene function analysis, which is why Gene Expression Databases, such as Gene Expression Omnibus (GEO), are essential (Edgar et al. 2002). However, GEO, which is a nonspecific database, does not hold any data from amaranth species (Edgar et al. 2002; Accessed 12 April, 2022).

There are plant-specific gene expression databases: DRASTIC (Button et al. 2006), PLANEX (Yim et al. 2013), and PLEXdb (Dash et al. 2012); however, no data is available for the *Amaranthus* genus. These databases hold data of genes regulated in response to abiotic and biotic stresses, as well genes important to the phytopathogens (Dash et al. 2012; Yim et al. 2013).

Although there is no data on amaranth in the gene expression databases, some authors have carried out research on the subject. Délano-Frier et al. (2011) were pioneers in attempting to increase the limited transcriptomic information on amaranth, a highly nutritious and stress-tolerant crop. Briefly, they performed large-scale transcriptome analysis from different tissues (leaves and stems) of *A. hypochondriacus,* in response to several biotic (herbivory and bacterial infection) and abiotic stresses (water and salt stresses), using 454 pyrosequencing (Délano-Frier et al. 2011). The comparative transcriptome analyses returned 8260 homologous sequences between *A. tuberculatus* and *A. hypochondriacus* transcriptomes, and 1971 DEG were identified in response to at least 1 of the treatments applied (Délano-Frier et al. 2011). This study greatly contributed to underpin the multiple stress resistance in plants, a desired trait for plant biotechnology (Délano-Frier et al. 2011).

The lack of expression data on amaranth in the main databases reinforces the need of increasing functional gene analysis research of this crop, with transcriptomic experiments, such as RNA-seq, with different tissues and conditions. The increasing amount of data is not the only limiting factor to improve the knowledge on amaranth gene expression: the data integration is likewise important, facilitating research. This knowledge will be useful to reveal genes and mechanisms related to agronomic traits of interest, such as increased nutritional value and resistant varieties.

6.2.3 Amaranth and Protein Databases

In addition, transcriptomic information, proteomic data, and metabolomic data are crucial for understanding how plants function. The information necessary to create new amaranth varieties with desirable nutritional traits is directly related to functional omics data. Therefore, the UniProt (Universal Protein Resource) (The UniProt Consortium 2018) is the main curated database of annotated protein sequences. There are 2090 protein records on UniProt related to *Amaranthus* genus, in which only 23 were curated, and are now available on Swiss-Prot (The UniProt Consortium 2018; Accessed 5 April, 2022). Rodríguez et al. reviewed the amaranth potential bioactive proteins from the UniProt database with potential to prevent chronic diseases (Montoya-Rodríguez et al. 2015).

In addition, protein 3D structures are very important to design food of high nutrition and benefits for health. The Protein Data Bank (PDB) archive is the only repository for protein (Altschul et al. 1997). A quick search in PDB using the term "viridiplantae" returns 7165 structures (Altschul et al. 1997, Accessed 24 June, 2022). Of those, 17 protein structures seem to be related to the *Amaranthus* genus (when the term "*Amaranthus*" was searched) (Altschul et al. 1997, Accessed 24 June, 2022). However, five of the retrieved matches were from heterologous expression systems, mainly *Zea mays* (Altschul et al. 1997, Accessed 24 June, 2022). Another option to increase amaranth 3D structures and function is using protein modeling tools, molecular dynamics, and molecular docking (Rasheed et al. 2020). Although bioinformatics is a growing science, only 9% of the structures from RSCB PDB (Research Collaboratory for Structural Bioinformatics Protein Data Bank) are from plants (Rasheed et al. 2020). Rasheed et al. (2020) claimed that the elucidation of seed storage protein structure and function is crucial to plant applications in the food industry. However, few studies regarding the structure of storage proteins have been performed. Therefore, amaranth structure prediction using bioinformatics might lead to the prediction of targets for enhancing the nutritional value of this plant (Rasheed et al. 2020).

Increasing proteomics research is of great importance to elucidate proteins with economic importance for this plant. For instance, the effect of an *A. cruentus* peptidome fraction on enzymes involved in the cholesterol biosynthesis was studied by Soares et al. (2015). This study showed that very small peptides (under 3 kDa) from amaranth might be involved in the hypocholesterolemic effect of this grain (Soares et al. 2015). These results show the importance of proteomics and peptidome's studies of amaranth to identify bioactive peptides with importance to health and nutrition.

In conclusion, with the arising data from HT NGS (high-throughput next-generation sequencing), several specific databases for genomic, transcriptomic, and proteomic data have been created, as we previously stated. However, the data is scattered across the platforms in many different formats, making it difficult to mine innovative knowledge. Integrative bioinformatics attempts to solve this problem by unifying this biological data (Quintans et al. 2022).

6.3 Amaranth and Comparative Genomics

Another powerful resource to analyze data from next-generation sequencing (NGS) is the Comparative Genome Databases (GGD) such as CoGe (Lyons and Freeling 2008) and Phytozome (Goodstein et al. 2012), which we will further discuss.

6.3.1 CoGe

The CoGe (Comparative Genomics research) holds data from 21,016 organisms and 56,091 genomes, where 23 genomes are from 4 species of the *Amaranthus* genus: *A. hybridus*, *A. hypochondriacus*, *A. palmeri*, and *A. tuberculatus* (Lyons and Freeling 2008; Accessed in June 24, 2022). The tools available for the comparative genome analyses at CoGe are CoGeBlast (Lyons et al. 2008), which performs an alignment of a query sequence with other genomes; SynMap (Haug-Baltzell et al. 2017), which identifies syntenic regions between genomic sequences from two organisms and creates a dotplot; SynMap3D, which plots the results from synteny analyses, from up to three genomes, in 3D (Haug-Baltzell et al. 2017); the SynFind tool, which identifies syntenic regions in whichever set of genomes; and GEvo, which is a genomic evolutionary tool analysis. Additionally, genomic features can be analyzed and visualized with tools such as FeatView and LoadExp+ (Grover et al. 2017; Lyons and Freeling 2008).

For instance, SynMap4.2 (Haug-Baltzell et al. 2017) has been used to infer synteny analysis between *A. hypochondriacus*, *B. vulgaris* (Ref Beet-1.1), and *Arabidopsis thaliana* (genome assembly TAIR10) revealing insights about amaranth evolution, such as the events of polyploidization, and a high correlation with other members of Amaranthaceae family (*B. vulgaris*) (Clouse et al. 2016; Lightfoot et al. 2017). Montgomery and collaborators (2020) also inferred evolution events (such as genomic duplication and translocation) among amaranth genomes by the comparison of *Amaranthus tuberculatus*, *Amaranthus hybridus*, and *Amaranthus palmeri* with *Amaranthus hypochondriacus*, using Sinmap2 (Haug-Baltzell et al. 2017). Besides, the genetic mapping performed in this study was used to authenticate the quality of the scaffolds and revealed a need to use different methods in order to obtain a higher quality genomic assembly for *A. tuberculatus* (Montgomery et al. 2020). Regarding the genomic composition, Sunil et al. (2014) used the SinMaP to compare amaranth genome with other plants, such as *S. lycopersicum*, *V. vinifera*, and *A. thaliana*, and found that *A. hypochondriacus* A-T (66%) content is very comparable within the studied plants, revealing common genomic features.

Transcriptomic comparative analyses are also possible using CoGe tools (Grover et al. 2017). For instance, the tools available are tRNAView, which searches for tRNAs in genomes or a given genomic sequence; the MatrixView which creates a dotPlot between two genomes and may be used to analyze genomic duplications (Nguyen et al. 2019); QuotaAlign, which analyzes and screens for whole syntenic blocks, based on expected coverage; GenomeMap, which is a visual map of the distribution of genomic features in the genome; GenomeView, which is a CoGe's

interactive genome browser (also known as EPIC-CoGe); and CodeOn, which is a tool for generating an amino acid usage table binned by the C-G content of each CDS (Grover et al. 2017). Regarding amaranth comparative transcriptome analysis, using SynMap (Haug-Baltzell et al. 2017), Sunil et al. (2014) have found that amaranth transcriptome is also rich in A-T content, which is also very comparable to other plant transcriptomes. The degree of synteny among genomes can also be used to infer the nutritional quality of plants, by revealing genomic features of interest (Ma et al. 2021).

6.3.2 Phytozome

There are databases and tools for comparing genes and genomes specific to plants, such as the Phytozome (Goodstein et al. 2012). Phytozome includes several genomic sequences from selected algae to land plants, allowing the researcher to infer their evolutionary stories from a molecular level (Goodstein et al. 2012).

Phytozome v13 (Goodstein et al. 2012) holds 274 assembled and annotated genomes, where only the *A. hypochondriacusv2.1* genome is available for the *Amaranthus* genus, with an estimated size of 466 Mb (Clouse et al. 2016; Lightfoot et al. 2017). This database is very useful to easily retrieve gene sequences from *A. hypochondriacus* that will support research with this plant. Beyond the dataset, Phytozome provides several tools to analyze and compare amaranth genome with others species, such as BLAST (a search tool based on sequence similarity) (Altschul et al. 1997), JBrowse (a web-tool for genome visualization and analysis) (Buels et al. 2016), and PhytoMine, which is an interface to visualize data from Phytozome.

7 Conclusion and Prospects

This work presented the recent advancement in amaranth research focusing on the genetic resources that can be leveraged to improve the crop. After decades of neglect, the crop was redomesticated and brought to the attention of farmers and researchers. Despite its exceptional nutritional quality, medicinal and industrial uses, climate resiliency, and ability to thrive in a broad range of climatic zones, the crop is still underutilized. Since the pioneering work of Jonathan Sauer in the 1950s, there has been slow progression in the development of taxonomic, biological, and molecular information on the crop. In recent years, reference genome and transcriptome information are developed, and several genes related to biotic and abiotic stresses, growth, and development were identified. Reverse genetics tools such as VIGs were developed for amaranths; however, there is still a lack of functional genomics work; therefore, numerous putative genes with potentially important agricultural traits are not functionally validated. Molecular breeding efforts especially utilizing multi-omics approaches are lacking. Most of the genomics and transcriptomics work involve domesticated species; there is also a need for high-quality "omics" information comparing wild-type, cultivated, and weedy amaranths; DNA fingerprints from the comparative studies aid not only in the discovery of genes and QTLs that are agronomically useful but also in high-throughput genotyping and understanding of

the complexity that exists in the *Amaranthus* genus. Taken together, all this omics information helps in the molecular improvement of the crop, which can add nutritional value to the agricultural systems for our future generations.

References

Adhikari S, Khatri-Chhetri U, Adhikary D (2021) Genome and transcriptome of Amaranth Species. In: Adhikary, D, Deyholos, MK, Délano-Frier, JP (eds) The Amaranth Genome. Compendium of Plant Genomes. Springer, Cham. https://doi.org/10.1007/978-3-030-72365-1_1

Adhikary D, Khatri-Chhetri U, Tymm FJM, Murch SJ, Deyholos MK (2019) A virus-induced gene-silencing system for functional genetics in a betalainic species, *Amaranthus tricolor* (Amaranthaceae). Appl Plant Sci 7(2). https://doi.org/10.1002/aps3.1221

Adhikary D, Khatri-Chhetri U, Slaski J (2020) Amaranth: an ancient and high-quality wholesome crop. In: Waisundara VY (ed) Nutritional value of Amaranth. IntechOpen. https://doi.org/10.5772/intechopen.88093

Ahmar S, Mahmood T, Fiaz S, Mora-Poblete F, Shafique MS, Chattha MS, Jung K-H (2021) Advantage of nanotechnology-based genome editing system and its application in crop improvement. Front Plant Sci:12. https://doi.org/10.3389/fpls.2021.663849

Akin-Idowu PE, Gbadegesin MA, Orkpeh U, Ibitoye DO, Odunola OA (2016) Characterization of grain Amaranth (*Amaranthus* spp.) germplasm in south West Nigeria using morphological, nutritional, and random amplified polymorphic DNA (RAPD) analysis. Resources 5(1). https://doi.org/10.3390/resources5010006

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17):3389–3402

Bai X, Ye X, Luo Y, Liu C, Wu Q (2021) Characterization of the first complete chloroplast genome of Amaranthus hybridus (Caryophyllales: Amaranthaceae) with phylogenetic implications. Mitochondrial DNA Part 6(11):3306–3308. https://doi.org/10.1080/23802359.2021.1994890

Baltensperger D, Weber L, Nelson L (1992) Registration of "Plainsman" grain amaranth. Crop Sci 32:1510–1511. https://doi.org/10.2135/cropsci1992.0011183X003200060047x

Berná G, Oliveras-López MJ, Jurado-Ruíz E, Tejedo J, Bedoya F, Soria B, Martín F (2014) Nutrigenetics and nutrigenomics insights into diabetes etiopathogenesis. Nutrients 6(11): 5338–5369. https://doi.org/10.3390/nu6115338

Böhm H, Rink E (1988) Chapter 26 – Betalains. In: Constabe LF, Vasil IK (eds) Phytochemicals in plant cell cultures. Academic, pp 449–463. https://doi.org/10.1016/B978-0-12-715005-5.50033-9

Bortesi L, Fischer R (2015) The CRISPR/Cas9 system for plant genome editing and beyond. Biotechnol Adv 33:41–52. https://doi.org/10.1016/j.biotechadv.2014.12.006

Buels R, Yao E, Diesh CM et al (2016) JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biol 17(1):66. https://doi.org/10.1186/s13059-016-0924-1

Button DK, Gartland KMA, Ball LD, Natanson L, Gartland JS, Lyon GD (2006) DRASTIC – INSIGHTS: querying information in a plant gene expression database. Nucleic Acids Res 34: D712–D716. https://doi.org/10.1093/nar/gkj136

CABI (2022) Invasive species compendium. www.cabi.org/isc

Chakraborty S, Chakraborty N, Agrawal L, Ghosh S, Narula K, Shekhar S, Naik PS, Pande PC, Chakrborti SK, Datta A (2010) Next-generation protein-rich potato expressing the seed protein gene *AmA1* is a result of proteome rebalancing in transgenic tuber. Proc Natl Acad Sci U S A 107(41):17533–17538. https://doi.org/10.1073/pnas.1006265107

Chandi A, Milla-Lewis SR, Jordan DL, York AC, Burton JD, Zuleta MC, Whitaker JR, Culpepper AS (2013) Use of AFLP markers to assess genetic diversity in palmer Amaranth (*Amaranthus palmeri)* populations from North Carolina and Georgia. Weed Sci 61(1):136–145. https://doi.org/10.1614/WS-D-12-00053.1. Cambridge Core

Chaney L, Mangelson R, Ramaraj T, Jellen EN, Maughan PJ (2016) The complete chloroplast genome sequences for four *Amaranthus* species (Amaranthaceae). Appl Plant Sci 4(9):1600063. https://doi.org/10.3732/apps.1600063

Clouse JW, Adhikary D, Page JT, Ramaraj T, Deyholos MK, Udall JA, Fairbanks DJ, Jellen EN, Maughan PJ (2016) The Amaranth genome: genome, transcriptome, and physical map assembly. Plant Genome 9(1). https://doi.org/10.3835/plantgenome2015.07.0062

Coelho LM, Silva PM, Martins JT, Pinheiro AC, Vicente AA (2018) Emerging opportunities in exploring the nutritional/functional value of amaranth. Food Funct 9(11):5499–5512. https://doi.org/10.1039/c8fo01422a

D'Amico S, Schoenlechner R (2017) Chapter 6 – Amaranth: its unique nutritional and health-promoting attributes. In: Taylor JRN, Awika JM (eds) Gluten-free ancient grains. Woodhead Publishing, pp 131–159. https://doi.org/10.1016/B978-0-08-100866-9.00006-6

Dar AA, Mahajan R, Sharma S (2019) Molecular markers for the characterization and conservation of plant genetic resources. Indian J Agric Sci 89(11):1755–1763

Dash S, Van Hemert J, Hong L, Wise RP, Dickerson JA (2012) PLEXdb: gene expression resources for plants and plant pathogens. Nucleic Acids Res 40:D1194–D1201. https://doi.org/10.1093/nar/gkr938

Das S (2016) Distribution and maintenance of Amaranth Germplasm Worldwide. In: Amaranthus: A promising crop of future. Springer, Singapore. https://doi.org/10.1007/978-981-10-1469-7_7

Datta SC (2021) Weed plant act as vaccine against plant and COVID-19 diseases: enriched agriculture health development socioeconomy sciences technology communication application. IJPSCR. https://doi.org/10.22377/ijpscr.v1i1.10

Délano-Frier JP, Avilés-Arnaut H, Casarrubias-Castillo K, Casique-Arroyo G, Castrillón-Arbeláez PA et al (2011) Transcriptomic analysis of grain amaranth (*Amaranthus hypochondriacus*) using 454 pyrosequencing: comparison with *A. tuberculatus*, expression profiling in stems and in response to biotic and abiotic stress. BMC Genomics 12:363. https://doi.org/10.1186/1471-2164-12-363

Dharajiya DT, Singh AK, Tiwari KK, Prajapati NN (2021) Genetic diversity in Amaranth and its close relatives. In: Adhikary D, Deyholos MK, Délano-Frier JP (eds) The Amaranth genome. Springer International Publishing, Cham, pp 81–96. https://doi.org/10.1007/978-3-030-72365-1_6

Eastman C, Zimmermann M (2000) The iodine deficiency disorders. [Updated 2018 Feb 6]. In: Feingold KR, Anawalt B, Boyce A et al (eds) Endotext [Internet]. MDText.com, Inc, South Dartmouth. https://www.ncbi.nlm.nih.gov/books/NBK285556/

Edgar R, Domrachev M, Lash AE (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30:207–210. https://doi.org/10.1093/nar/30.1.207

Emmanuel OC, Babalola OO (2021) Amaranth production and consumption in South Africa: the challenges of sustainability for food and nutrition security. Int J Agric Sustain 20:1–12. https://doi.org/10.1080/14735903.2021.1940729

Floret C, Monnet A-F, Micard V, Walrand S, Michon C (2021) Replacement of animal proteins in food: how to take advantage of nutritional and gelling properties of alternative protein sources. Critl Rev Food Sci Nutr 63:1–27. https://doi.org/10.1080/10408398.2021.1956426

Garg M, Sharma N, Sharma S, Kapoor P, Kumar A, Chunduri V, Arora P (2018) Biofortified crops generated by breeding, agronomy, and transgenic approaches are improving lives of millions of people around the world. Front Nutr 5. https://doi.org/10.3389/fnut.2018.00012

Gebreil SY, Ali MIK, Mousa EAM (2020) Utilization of Amaranth flour in preparation of high nutritional value bakery products. Food Nutr Sci 11(5):19

Gerrano A, Jansen van Rensburg W, Mavengahama S, Bairu M, Venter S, Adebola P (2017) Qualitative morphological diversity of *Amaranthus* species. J Trop Agric 55:12–20

Gonçalves-Dias J, Stetter MG (2021) PopAmaranth: a population genetic genomebrowser for grain amaranths and their wild relatives. Genes Genomes Genet 11(7). https://doi.org/10.1093/g3journal/jkab103

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD et al (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178–D1186. https://doi.org/10.1093/nar/gkr944

Govindaraj M, Vetriventhan M, Srinivasan M (2015) Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. Genet Res Int 2015:431487. https://doi.org/10.1155/2015/431487

Greene CS, Tan J, Ung M, Moore JH, Cheng C (2014) Big data bioinformatics. J Cell Physiol 229: 1896–1900. https://doi.org/10.1002/jcp.24662

Grover JW, Bomhoff M, Davey S, Gregory BD, Mosher RA, Lyons E (2017) CoGe LoadExp+: a web-based suite that integrates next-generation sequencing data analysis workflows and visualization. Plant Direct 1(2). https://doi.org/10.1002/pld3.8

Grützner R, Schubert R, Horn C, Yang C, Vogt T, Marillonnet S (2021) Engineering Betalain biosynthesis in tomato for high level Betanin production in fruits. Front Plant Sci 12. https://doi.org/10.3389/fpls.2021.682443

Gupta MK, Donde R, Sabarinathan S, Gouda G, Dash GK, Pati P, Behera L (2021) Microsatellite markers from whole genome and transcriptomic sequences. In: Bioinformatics in rice research. Springer, Singapore, pp 387–412

Haug-Baltzell A, Stephens SA, Davey S, Scheidegger CE, Lyons E (2017) SynMap2 and SynMap3D: web-based whole-genome synteny browsers. Bioinformatics 33(14):2197–2198. https://doi.org/10.1093/bioinformatics/btx144

Holesh J, Aslam S, Martin A (2022) Physiology, carbohydrates. StatPearls [Internet]. https://www.ncbi.nlm.nih.gov/books/NBK459280/

Hollie A, Grabitske MS, Joanne LS (2009) Gastrointestinal effects of low-digestible carbohydrates. Crit Rev Food Sci Nutr 49(4):327–360. https://doi.org/10.1080/10408390802067126

Hong S-Y, Cheon K-S, Yoo K-O, Lee H-O, Mekapogu M, Cho K-S (2019) Comparative analysis of the complete chloroplast genome sequences of three Amaranthus species. Plant Genet Resour Characteriz Utiliz 17(3):245–254. https://doi.org/10.1017/S1479262118000485. Cambridge Core

Hoshikawa K, Lin Y-P, Schafleitner R, Shirasawa K, Isobe S, Nguyen DC, Ohsawa R, Yoshioka Y (2023) Genetic diversity analysis and core collection construction for Amaranthus tricolor germplasm based on genome-wide single-nucleotide polymorphisms. Sci Hortic 307:111428. https://doi.org/10.1016/j.scienta.2022.111428

Ibrahim, N'I, Fairus S, Zulfarina MS, Naina Mohamed I (2020) The efficacy of squalene in cardiovascular disease risk-a systematic review. Nutrients 12(2). https://doi.org/10.3390/nu12020414

Ivanova S, Delattre C, Karcheva-Bahchevanska D, Benbasat N, Nalbantova V, Ivanov K (2021) Plant-based diet as a strategy for weight control. Foods 10(12). https://doi.org/10.3390/foods10123052

Jamalluddin N, Massawe FJ, Mayes S, Ho WK, Symonds RC (2022) Genetic diversity analysis and marker-trait associations in Amaranthus species. PLoS One 17(5):e0267752. https://doi.org/10.1371/journal.pone.0267752

Jiang S-Y, Ma A, Xie L, Ramachandran S (2016) Improving protein content and quality by overexpressing artificially synthetic fusion proteins with high lysine and threonine constituent in rice plants. Sci Rep 6(1):34427. https://doi.org/10.1038/srep34427

Jimoh MO, Afolayan AJ, Lewu FB (2018) Suitability of Amaranthus species for alleviating human dietary deficiencies. S Afr J Bot 115:65–73. https://doi.org/10.1016/j.sajb.2018.01.004

Joshi DC, Sood S, Hosahatti R, Kant L, Pattanayak A, Kumar A, Yadav D, Stetter MG (2018) From zero to hero: the past, present and future of grain amaranth breeding. Theor Appl Genet 131(9): 1807–1823. https://doi.org/10.1007/s00122-018-3138-y

Joshi DC, Meena RP, Chandora R (2021) Chapter 3 – Genetic resources: collection, characterization, conservation, and documentation. In: Singh M, Sood S (eds) Millets and pseudo cereals. Woodhead Publishing, pp 19–31. https://doi.org/10.1016/B978-0-12-820089-6.00003-3

Kamble S, Gaikwad D (2021) Studies comparative studies on morphology of selected six species of the genus *Amaranthus* L. from India. Int J Bot 6:637–640

Kim S-K, Karadeniz F (2012) Biological importance and applications of squalene and squalane. Adv Food Nutr Res 65:223–233. https://doi.org/10.1016/B978-0-12-416003-3.00014-7

Kumar Pandey V, Jatt B, Pandey R, Varun Kumar Pandey C (2019) Marker assisted genetic diversity among *Amaranthus* species through RAPD. J Pharmacogn Phytochem (3):83–89

Liang Z, Zhang K, Chen K, Gao C (2014) Targeted mutagenesis in Zea mays using TALENs and the CRISPR/Cas system. J Genet Genomics 41(2):63–68. https://doi.org/10.1016/j.jgg.2013.12.001

Lightfoot DJ, Jarvis DE, Ramaraj T, Lee R, Jellen EN, Maughan PJ (2017) Single-molecule sequencing and Hi-C-based proximity-guided assembly of amaranth (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. BMC Biol 15(1):74. https://doi.org/10.1186/s12915-017-0412-4

Lin YP, Wu TH, Chan YK, van Zonneveld M, Schafleitner R (2022) *De novo* SNP calling reveals the genetic differentiation and morphological divergence in genus *Amaranthus*. Plant Genome 15(2):1–19. https://doi.org/10.1002/tpg2.20206

Lopez MJ, Mohiuddin SS (2021) Biochemistry, essential amino acids. StatPearls Publishing, Treasure Island. http://europepmc.org/books/NBK557845

Lyons E, Freeling M (2008) How to usefully compare homologous plant genes and chromosomes as DNA sequences. Plant J 53(4):661–673. https://doi.org/10.1111/j.1365-313X.2007.03326.x

Lyons E, Pedersen B, Kane J et al (2008) Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with Rosids. Plant Physiol 148(4):1772–1781. https://doi.org/10.1104/pp.108.124867

Ma X, Vaistij FE, Li Y, Jansen van Rensburg WS, Harvey S et al (2021) A chromosome- level *Amaranthus cruentus* genome assembly highlights gene family evolution and biosynthetic gene clusters that may underpin the nutritional value of this traditional crop. Plant J 107(2):613–628. https://doi.org/10.1111/tpj.15298

Malik KA, Maqbool A (2020) Transgenic crops for biofortification. Front Sustain Food Syst 4. https://doi.org/10.3389/fsufs.2020.571402

Maughan PJ, Smith SM, Fairbanks DJ, Jellen EN (2011) Development, characterization, and linkage mapping of single nucleotide polymorphisms in the grain amaranths (*Amaranthus* sp.). Plant Genome 4(1). https://doi.org/10.3835/plantgenome2010.12.0027

Montgomery JS, Giacomini D, Waithaka B et al (2020) Draft genomes of *Amaranthus tuberculatus*, *Amaranthus hybridus*, and *Amaranthus palmeri*. Genome Biol Evol 12(11):1988–1993. https://doi.org/10.1093/gbe/evaa177

Montoya-Rodríguez A, Gómez-Favela MA, Reyes-Moreno C, Milán-Carrillo J, González de Mejía E (2015) Identification of bioactive peptide sequences from amaranth (*Amaranthus hypochondriacus*) seed proteins and their potential role in the prevention of chronic diseases. Compr Rev Food Sci Food Saf 14(2):139–158. https://doi.org/10.1111/1541-4337.12125

Naik BJ, Shimoga G, Kim S-C, Manjulatha M, Subramanyam Reddy C, Palem RR, Kumar M, Kim S-Y, Lee S-H (2022) CRISPR/Cas9 and nanotechnology pertinence in agricultural crop refinement. Front Plant Sci 13. https://doi.org/10.3389/fpls.2022.843575

Nardo AE, Suárez S, Quiroga AV, Añón MC (2020) Amaranth as a source of antihypertensive peptides. Front Plant Sci 11. https://doi.org/10.3389/fpls.2020.578631

NCBI Resource Coordinators (2016) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 44(1):D7–D19. https://doi.org/10.1093/nar/gkv1290

Nguyen DC, Tran DS, Tran TTH, Ohsawa R, Yoshioka Y (2019) Genetic diversity of leafy amaranth (*Amaranthus tricolor* l.) resources in Vietnam. Breed Sci 69(4):640–650. https://doi.org/10.1270/jsbbs.19050

Ochieng J, Schreinemachers P, Ogada M, Dinssa FF, Barnos W, Mndiga H (2019) Adoption of improved amaranth varieties and good agricultural practices in East Africa. Land Use Policy 83:187–194. https://doi.org/10.1016/j.landusepol.2019.02.002

Oduwaye OA, Ayo-Vaughan MA, Porbeni JBO, Oyelakin OO (2019) Genetic diversity in amaranth (*Amaranthus* spp.) based on phenotypic and RAPD markers. Niger J Biotechnol 36(1):62. https://doi.org/10.4314/njb.v36i1.9

Packard EE, Quintans ILA d CR, Adhikary D (2021) Genetics of betalain pigments in amaranth species. In: Adhikary D, Deyholos MK, Délano-Frier JP (eds) The Amaranth genome. Springer International Publishing, Cham, pp 35–54. https://doi.org/10.1007/978-3-030-72365-1_4

Palmeros-Suárez PA, Casarrubias-Castillo K, Massange-Sánchez JA (2021) Amaranth transcription factors in response to biotic and abiotic stresses. In: Adhikary D, Deyholos MK, Délano-Frier JP (eds) The Amaranth genome. Springer International Publishing, Cham, pp 167–181. https://doi.org/10.1007/978-3-030-72365-1_10

Pulvento C, Sellami MH, Lavini A (2021) Yield and quality of Amaranthus hypochondriacus grain amaranth under drought and salinity at various phenological stages in southern Italy. J Sci Food Agric. https://doi.org/10.1002/jsfa.11088

Quintans ILA d CR, de Araújo JVA, Rocha LNM, de Andrade AEB, do Rêgo TG, Deyholos MK (2022) An overview of databases and bioinformatics tools for plant antimicrobial peptides. Curr Protein Pept Sci 23(1):6–19. https://doi.org/10.2174/1389203723666211222170342

Rasheed F, Markgren J, Hedenqvist M, Johansson E (2020) Modeling to understand plant protein structure-function relationships-implications for seed storage proteins. Molecules (Basel, Switzerland) 25(4):873. https://doi.org/10.3390/molecules25040873

Ryan E, Galvin K, O'Connor TP, Maguire AR, O'Brien NM (2007) Phytosterol, squalene, tocopherol content and fatty acid profile of selected seeds, grains, and legumes. Plant Foods Hum Nutr 62(3):85–91. https://doi.org/10.1007/s11130-007-0046-8

Sarker U, Oba S, Daramy MA (2020) Nutrients, minerals, antioxidant pigments and phytochemicals, and antioxidant capacity of the leaves of stem amaranth. Sci Rep 10(1):3892. https://doi.org/10.1038/s41598-020-60252-7

Sattar N, McInnes IB, McMurray JJV (2020) Obesity is a risk factor for severe COVID-19 infection. Circulation 142(1):4–6. https://doi.org/10.1161/CIRCULATIONAHA.120.047659

Sauer J (1957) Recent migration and evolution of the dioecious amaranths. Evolution 11(1):11–31. https://doi.org/10.2307/2405808

Schafleitner R, Lin Y, Dinssa FF et al (2022) The World Vegetable Center Amaranthus germplasm collection: core collection development and evaluation of agronomic and nutritional traits. Crop Sci 62:1173. https://doi.org/10.1002/csc2.20715

Showemimo F, Soyombo MA, Amira JO, Porbeni JBO (2021) Traits selection criteria for genetic improvement of grain and leafy amaranth (*Amaranthus* spp) using principal component analysis. Egypt J Agric Res 99(2):170–179. https://doi.org/10.21608/ejar.2021.65946.1089

Sisti MS, Scilingo A, Añón MC (2019) Effect of the incorporation of Amaranth (*Amaranthus mantegazzianus*) into fat- and cholesterol-rich diets for Wistar rats. J Food Sci 84(11):3075–3082. https://doi.org/10.1111/1750-3841.14810

Soares RAM, Mendonça S, de Castro LÍA, Menezes ACCCC, Arêas JAG (2015) Major peptides from amaranth (*Amaranthus cruentus*) protein inhibit HMG-CoA reductase activity. Int J Mol Sci 16(2):4150–4160. https://doi.org/10.3390/ijms16024150

Soriano-García M, Arias-Olguín I, Montes JPC et al (2018) Nutritional functional value and therapeutic utilization of Amaranth. J Anal Pharm Res 7(5):596–600. https://doi.org/10.15406/japlr.2018.07.00288

Stetter MG, Zeitler L, Steinhaus A, Kroener K, Biljecki M, Schmid KJ (2016) Crossing methods and cultivation conditions for rapid production of segregating populations in three grain *Amaranth* species. Front Plant Sci 7. https://doi.org/10.3389/fpls.2016.00816

Stetter MG, Vidal-Villarejo M, Schmid KJ (2020) Parallel seed color adaptation during multiple domestication attempts of an ancient new world grain. Mol Biol Evol 37(5):1407–1419. https://doi.org/10.1093/molbev/msz304

Sunil M, Hariharan AK, Nayak S, Gupta S, Nambisan SR, Gupta RP, Panda B, Choudhary B, Srinivasan S (2014) The draft genome and transcriptome of *Amaranthus hypochondriacus*: a C4 dicot producing high-lysine edible pseudo-cereal. DNA Res 21(6):585–602. https://doi.org/10.1093/dnares/dsu021

Tabio-García D, Paraguay-Delgado F, Sánchez-Madrigal MÁ, Quintero-Ramos A, Espinoza-Hicks JC, Meléndez-Pizarro CO, Ruiz-Gutiérrez MG, Espitia-Rangel E (2021) Optimisation of the ultrasound-assisted extraction of betalains and polyphenols from *Amaranthus hypochondriacus* var. Nutrisol. Ultrason Sonochem 77:105680. https://doi.org/10.1016/j.ultsonch.2021.105680

Tamás C, Kisgyörgy BN, Rakszegi M, Wilkinson MD, Yang M-S, Láng L, Tamás L, Bedő Z (2009) Transgenic approach to improve wheat (*Triticum aestivum* L.) nutritional quality. Plant Cell Rep 28(7):1085–1094. https://doi.org/10.1007/s00299-009-0716-0

Thapa R, Blair MW (2018) Morphological assessment of cultivated and wild Amaranth species diversity. Agronomy 8(11). https://doi.org/10.3390/agronomy8110272

Thapa R, Edwards M, Blair MW (2021) Relationship of cultivated grain Amaranth species and wild relative accessions. Genes 12(2). https://doi.org/10.3390/genes12121849

The UniProt Consortium (2018) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47(D1):D506–D515. https://doi.org/10.1093/nar/gky1049

Tiwari KK, Thakkar NJ, Dharajiya DT, Bhilocha HL, Barvaliya PP, Galvadiya BP, Prajapati NN, Patel MP, Solanki SD (2021) Genome-wide microsatellites in amaranth: development, characterization, and cross-species transferability. 3 Biotech 11(9):395. https://doi.org/10.1007/s13205-021-02930-5

Trucco F, Tranel PJ (2011) Amaranthus. Springer, Berlin/Heidelberg. https://doi.org/10.1007/978-3-642-20450-0_2

Valdes AM, Walter J, Segal E, Spector TD (2018) Role of the gut microbiota in nutrition and health. BMJ 361:k2179. https://doi.org/10.1136/bmj.k2179

Varshney RK, Bohra A, Yu J, Graner A, Zhang Q, Sorrells ME (2021) Designing future crops: genomics-assisted breeding comes of age. Trends Plant Sci 26(6):631–649. https://doi.org/10.1016/j.tplants.2021.03.010

Wada N, Ueta R, Osakabe Y, Osakabe K (2020) Precision genome editing in plants: State-of-the-art in CRISPR/Cas9-based genome engineering. BMC Plant Biol 20(1):234. https://doi.org/10.1186/s12870-020-02385-5

Webb P (2021) The 2020 Nobel Peace Prize rewards the persistent vision of a world without hunger, famine, or malnutrition. Am J Clin Nutr 113(2):290–293. https://doi.org/10.1093/ajcn/nqaa355

Wenefrida I, Utomo HS, Linscombe SD (2013) Mutational breeding and genetic engineering in the development of high grain protein content. J Agric Food Chem 61:11702–11710. https://doi.org/10.1021/jf4016812

Wolter F, Schindele P, Puchta H (2019) Plant breeding at the speed of light: the power of CRISPR/Cas to generate directed genetic diversity at multiple sites. BMC Plant Biol 19(1):176. https://doi.org/10.1186/s12870-019-1775-1

Wu X, Blair MW (2017) Diversity in grain amaranths and relatives distinguished by genotyping by sequencing (GBS). Front Plant Sci 8. https://doi.org/10.3389/fpls.2017.01960

Yim WC, Yu Y, Song K, Jang CS, Lee B-M (2013) PLANEX: the plant co-expression database. BMC Plant Biol 13:83. https://doi.org/10.1186/1471-2229-13-83

Żukiewicz-Sobczak W, Wróblewska P, Zwoliński J, Chmielewska-Badora J, Adamczuk P, Krasowska E, Zagórski J, Oniszczuk A, Piątek J, Silny W (2014) Obesity and poverty paradox in developed countries. Ann Agric Environ Med 21(3):590–594. https://doi.org/10.5604/12321966.1120608

Cucumber (*Cucumis sativus* L.): Genetic Improvement for Nutraceutical Traits

Ashutosh Rai, Vishal Chugh, and Sudhakar Pandey

Contents

Abstract

Among the cucurbits, cucumber (*Cucumis sativus* L.) is a valuable vegetable crop cultivated for its immature fruits. The cucumber is one of the oldest cultivated vegetable crops. It has been known in history for over 5,000 years and probably originated in India. A lot of diversity exists with respect to shape, size, and color of the fruits which contain 0.4% protein, 2.5% carbohydrates, 1.5 mg iron, and 2 mg of vitamin C per 100 g of fresh weight. Fruits are good for

A. Rai · V. Chugh
College of Horticulture, Banda University of Agriculture and Technology, Banda, India
e-mail: ashutoshrai@buat.edu.in; vishalchugh@buat.edu.in

S. Pandey (✉)
ICAR-Indian Institute of Vegetable Research, Varanasi, India

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_57

people suffering from constipation, jaundice, and indigestion. The quality of cucumber depends on its total soluble solids, fruit firmness, desired fruit size, and other phytochemicals like antioxidant capacity, phenols, and vitamin C content. As *Ayurvedic* and conventional remedy, cucumber is used for various skin-related problems, including inflammation under the eyes, and sunburn, and is assumed to have cooling, curative, comforting, emollient, lenitive, anti-itching effect to irritated skin, and extended cosmetic effects. Due to the presence of numerous active constituents distributed throughout the plant parts, including vitamins, minerals, amino acids, phytosterols, phenolic acids, fatty acids, and curcurbitacin, cucumber exhibits a variety of pharmacological properties. Cucumber fruit that have an orange colored endocarp or mesocarp have high β-carotene content. Carotenoids play vital roles in human nutrition as potent precursor for various vitamins for biosynthetic pathways. Cucurbitacins, the bitter triterpenoid compounds found in cucurbits, are toxic to most of the organisms but are able to attract specialized insects. The expression of cucrbitacin depends on a Mendelian gene known as "*Bi.*" The enzyme oxidosqualene cyclase (OSC) catalyzes the biosynthesis of responsible triterpene carbon skeleton in fruits and plants. An oxidosqualene cyclase (OSC) gene in squash (*Cucurbita pepo* L.) known as cucurbitadienol synthase (CPQ) is the first enzyme of biosynthetic pathway of cucurbitacin. The mapping of fruit quality-related quantitative trait loci and metabolic pathway studies are enabling researchers to enhance quality traits.

Keywords

Cucumber · Nutraceuticals · Cosmetic · Antimicrobial activity · Fruit quality

1 Introduction

Among the cucurbits, cucumber (*Cucumis sativus* L.) is A valuable vegetable crop cultivated for its immature fruits and is one of the oldest cultivated vegetable crops. It has been known in history for over 5,000 years and probably originated in India (Whitaker & Davis 1962). Cucumber was domesticated about 3000 years ago; total cultivated area under cucumber and gherkins is 26,500 ha with 168,000 tones production and 7.5 t/ha productivity in India (FAOSTAT 2018). It is considered as an important salad crop grown both in north and lower as well as higher hills in India (Pandey & Kujur 2022) (Fig. 1).

A lot of diversity exists with respect to shape, size, and color of the fruits which contains 0.4% protein, 2.5% carbohydrates, 1.5 mg iron, and 2 mg of vitamin C per 100 g of fresh weight (Pandey et al. 2018). Fruits are good for people suffering from constipation, jaundice, and indigestion. The quality of cucumber depends on its total soluble solids, fruit firmness, desired fruit size, and other phytochemicals like antioxidant capacity, phenols, and vitamin C content.

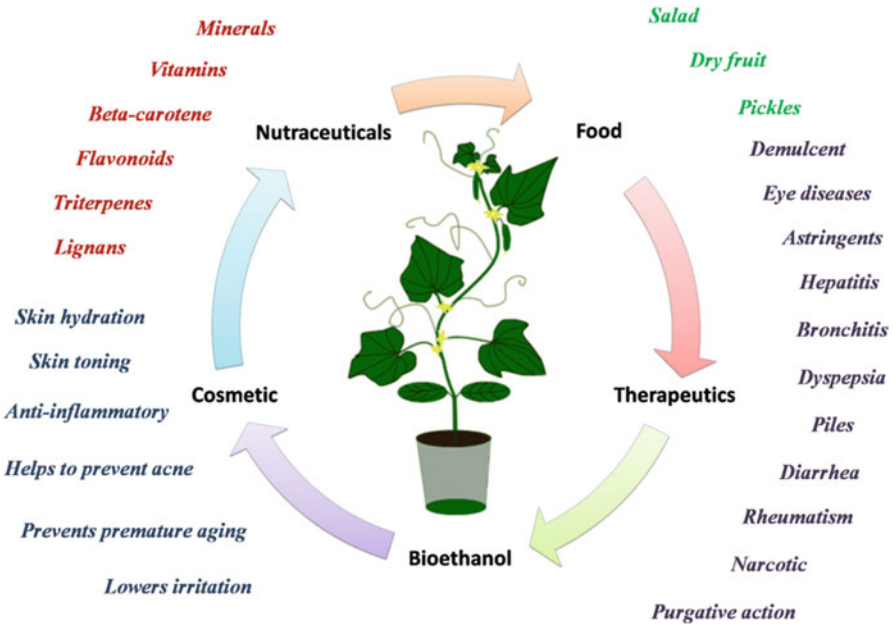


Fig. 1 Diverse sectors and uses of cucumber

2 Cucumber for Culinary Purposes

Most of the human diseases like cancer, inflammation, heart-related disease, diabetes, and autoimmune diseases as well as neurodegenerative diseases which affect neurotransmitter levels resulting various mental disorders are resultant of oxidative stress or disturbance in the cellular redox balance. The cell proliferation is promoted with more oxidizing conditions. Diet plays a crucial role including various phytochemicals maintaining human health. Cucumber is rich in water content and very low in calories. Cucumber is considered as vegetable crop rich in phytochemicals and polyphenolics. Major groups of dietary antioxidants-phytochemicals found in cucumber are vitamin C (ascorbic acid), folic acid, phenolic compounds (phenolic acids such as cinnamic acid and polyphenols such as flavonoids), and Phytocassane D; Cucurbitacin E; Cucurbitacin I; Nomilin; terpenes such as Limonin; and isoprenoids (vitamin E tocopherols and carotenoids). The exocarp of cucumber is also a rich source of different minerals such as calcium, magnesium, iron, potassium, phosphorus, sodium, manganese, zinc, and sulfur. These amino acids are also found in cucumber in their higher to lover concentrations as glutamine, oxoproline, alanine, glycine, citrulline, leucine, isoleucine, valine, tyrosine, serine, glutamic acid, gaba, aspartic acid, proline, phenylalanine, threonine, histidine, ornithine, lysine, methionine, 4-aminobutyric acid, beta-alanine, and homoserine (Zhao et al. 2016). The seeds of cucumber also contain

considerable quantity of minerals having mainly magnesium (8.5%), calcium (2.0%), manganese (66.3 ppm), sodium (79.4 ppm), and iron (21.4 ppm) with small quantity of zinc (11.7 ppm), copper (3.3 ppm), and potassium (5.1%) (Abiodun and Adeleke 2010). Cucumber is well known for its anticarcinogenic, antioxidant, antielastase, antihyaluronidase, hypolipidemic, antihyperglycemic, anti-inflammatory, diuretic, antimicrobial, amylolytic, and analgesic properties. With strong nutty flavor, cucumber seeds are used as dry fruit as well as to extract oil. Seed kernel of *C. sativus* contains proteins like globulins, albumins, and glyoxysomal enzymes in peroxisomes such as isocitrate lyase, malate synthase, citrate synthase, malate dehydrogenase, crotonase, and catalase (Köller et al. 1979).

Tender and fresh fruits of cucumber are consumed as salads and desserts and as cooked vegetables known as slicers, were as it is also preferred as preserved in form of pickles from cultivars know as picklers. The whitish to greenish mesocarp of cultivated cucumber has been of culinary importance for consumer preference worldwide. Diverse types of cucumber are commercially grown worldwide, together with American food processing and fresh market industries, Oriental trellis (burpless) varieties, European gherkin and glasshouse variety, the Mideast Beit Alpha type, and the German Schalgurken (Staub et al. 2008).

Tender fruits of particular small-fruited cultivars known as 'gherkin' are most excellently used pickled in vinegar. Cucumber fruits are also preserved in high concentration of salt. Cucumber in Korea as Oi sobagi kimchi and oiji as pickled, whereas in Egypt, brine soaking is common, known as torshi khiar. In many Asian countries, the seed kernels are dehusked and used in confectionary. The oil extracted from cucumber seeds is used for salad dressing in French cooking (Lim 2012).

3 Cosmetic Properties

As ayurvedic and conventional remedy, cucumber is used for various skin-related problems, including inflammation under the eyes and sunburn, and is assumed to have cooling, curative, comforting, emollient, lenitive, anti-itching effect on irritated skin, and extended cosmetic effects (Sotiroudis et al. 2010). Cucumber fruits are extensively used as skin-cleansing agent and are an integral part of various skin-whitening- and -softening-related cosmetic formulations. In addition, the cucumber fruit juice is utilized for a number of cosmetic formulations (Grieve 1998; Chiej 1984).

4 Nutraceuticals and Therapeutic Properties

Due to the presence of numerous active constituents distributed throughout the plant parts, including vitamins, minerals, amino acids, phytosterols, phenolic acids, fatty acids, and curcurbitacin, cucumber exhibits a variety of pharmacological properties. Cucumber fruit that have an orange-colored endocarp or mesocarp have high β-carotene content. Carotenoids play vital roles in human nutrition as potent precursor for various vitamins for biosynthetic pathways.

4.1 Antioxidant Activity

The cucumber fruit and seeds contain various antioxidants and secondary metabolites such as terpenoids, flavonoids, cardiac glycoside, tannins, phenols, and abundance of carbohydrates. The sturdy antioxidant activity and phytochemical analysis of ethanolic seeds extracts of cucumber have been performed by various workers (Begum et al. 2018). Methanolic extracts of cucumber pulp have also been reported to show powerful antioxidant activity through DPPH assay owing to its high amount of phenolics (Sotiroudis et al. 2010). Agarwal et al. (2012) reported good antioxidant activity in the aqueous solvent extracts from peel of the cucumber based on phospho molybdenum assay and established that this activity is due to the polyphenols content. In another study, cucumber fruit extract was tested for its ability to scavenge free radicals and functions as an analgesic. The extract was tested for analgesic effects at doses of 250 and 500 mg/kg and *in vitro* antioxidant investigations at 250 and 500 g/ml, respectively. Ascorbic acid and BHA (butylated hydroxyl anisole) were tested for their ability to scavenge free radicals, whereas diclofenac sodium (50 mg/kg) was used to examine their analgesic effects. The greatest antioxidant and analgesic effects of the cucumber fruit extract were observed at 500 g/ml and 500 mg/kg, respectively. According to early phytochemical screening, the extract contains flavonoids and tannins, which may be the cause of the analgesic and free radical-scavenging properties (Kumar et al. 2010).

4.2 Antimicrobial Activity

Three antimicrobial sphingolipids, first (2S,3S,4R,10E)-2-[(2′R)-2-hydroxy-tetracosanoyl amino]-1,3,4-octadecanetriol-10-ene, second 1-O-β-D-glucopyranosyl, and third soyacerebroside-I, have been identified in crude methanolic homogenate of cucumber stems. Cucumber extract exhibited antifungal and antibacterial activity against microorganisms, including four fungal and three bacterial species (Tang et al. 2010). Antimicrobial activity in the ethanolic extracts of seeds of cucumber has also been reported against *Staphylococcus aureus*, *Salmonella typhi, Acremonium, Verticellium, Pythium,* and *Tricoderma* spp. using disc diffusion assay (Begum et al. 2018). Antimicrobial activity against six Gram negative and Gram positive bacterial strains have been reported along with three human-pathogen fungi (*Candida albicans, C. tropicalis,* and *C. glabrata*) in methanolic and dichloromethane extracts of the fleshy pericarp and the exocarp of cucumber fruit (Sotiroudis et al. 2010).

4.3 Wound-Healing Activity

The aqueous extract cream formulation containing soft white paraffin as base in 2.5%, 5%, and 10% w/w of cucumber fruit have been reported for its ameliorative effects on 300 mm wide and 2 mm deep wounds in rats (Patil et al. 2011). Significant

Table 1 Nutraceutical compounds in different plant parts of cucumber and their use

Parts of plant	Therapeutic and medicinal uses	Reference
Fruit	Astringents, bronchitis, hepatitis, dyspepsia, diarrhea, piles, cough hoarseness of voice, asthma, and eye diseases	Mallik et al. (2013), Sahu and Sahu (2015)
Seed oil	Skin care properties like soothing, moisturizing, and protection from UV rays	Abiodun and Adeleke (2010)
Kernel of fruit	Rheumatism, narcotic, and purgative action	Campbell (2008), Mallik et al. (2013)
Pericarp of fruit	Dropsy, dysenteric-diarrhea, piles, and leprosy	Gopalakrishnan et al. (2014), Sahu and Sahu (2015)
Gum of bark	Demulcent, purgative	Arya et al. (2012), Mallik et al. (2013), Sahu and Sahu (2015)

decrease of $P < 0.05$, $P < 0.001$, and $P < 0.001$ was observed for wound area, epithelization period, and scar width, respectively. The faster wound contraction compared with control samples indicated that cucumber's wound-healing abilities are attributable to its antioxidant and scavenging properties due to various flavanoids and secondary metabolites. Phytofabrication has been achieved and found significantly wound healing potential in rat model. The ointment was obtained by encapsulated metallic silver-encapsulated bioactive molecules in metallic silver nanoparticles using callus (*CAgNPs*) and leaf extracts (*LEAgNPs*) of cucumber (Venkatachalam et al. 2015) (Table 1).

5 Genome Structure and Fruit Quality-Related Genes in Cucumber

Cucumis hystrix Chakr. is a wild species native to Asia whose fruits possess a flavor typical of cultivated cucumber (Chen et al. 1994). Although the literature indicates that *C. hystrix* is $2n = 2x = 14$ (Dane 1991), no definitive cytogenetic analysis has been made of this species, and its taxonomic placement has been based entirely on morphological characteristics (Kirkbride 1993). Isozyme analysis of *C. hystrix*, *C. sativus,* and *C. melo* led to the theory that a triangular phylogenetic relationship exists among these species (Chen et al. 1995).

Cucumis, a genus of twining, tendril-bearing plants in the family Cucurbitaceae, can be divided into two designated subgenera as *Cucumis* ($2n = 2x = 14$ and 24) and *Melo* ($2n = 2x = 24$) forming five number of cross-sterile species groups (Jeffrey, 1980). Furthermore, the subgenus *Cucumis* has been separated as Sino-Himalayan species, including *C. sativus* ($2n = 2x = 14$) and *C. hystrix* Chakr. ($2n = 2x = 24$). *C. sativus* is the most commercially utilized group including var. *sativus*, the cultivated cucumber, and the wild, naturally occurring var. *hardwickii* (R.) Alef. (Kirkbride 1993).

6 Biosynthesis of Phytochemicals in Cucumber

Cucurbitacins, the bitter triterpenoid compounds found in cucurbits, are toxic to most of the organisms but are able to attract specialized insects. The expression of cucrbitacin depends on a Mendelian gene known as "*Bi.*" The enzyme oxidosqualene cyclase (OSC) catalyzes the biosynthesis of responsible triterpene carbon skeleton in fruits and plants. An oxidosqualene cyclase (OSC) gene in squash (*Cucurbita pepo* L.) known as cucurbitadienol synthase (CPQ) is the first enzyme of the biosynthetic pathway of cucurbitacin (Shibuya et al. 2004).

Four oxidosqualene cyclase genes have been previously identified responsible for cucurbitacin synthesis; the ortholog of CPQ, *Csa008595,* has been recognized in cucumber which resides in a genetic interval that defines the *Bi* gene. It was also deduced that *Csa008595* constitutes a cluster of genes, which include one acyltransferase-encoding gene (*Csa008594*) and two cytochrome P450–encoding genes (*Csa008596* and *Csa008597*) (Huang et al. 2009). Authors have also reported that abovementioned three genes (*Csa008594*, *Csa008595,* and *Csa008597*) are expressed simultaneously strongly in cucumber leaf tissue in a similar fashion to that of the operon like gene cluster involved in thalianol biosynthesis in Arabidopsis (Field and Osbourn 2008). This gene set may consequently catalyze the stepwise formation of cucurbitacin in cucumber. Cucumber is considered as model system for studying sex expression in plants. Ethylene promotes the femaleness and is thought as the sex hormone of the cucumber (Rudich et al. 1972). Huang et al. (2009) also reported 137 genes in cucumber related to ethylene biosynthetic and signaling pathways. A crucial regulating enzyme in the ethylene biosynthesis pathway is 1-aminocyclopropane-1-carboxylate synthase (ACS), which is encoded by the melon gene *Cm-ACS7* (Boualem et al. 2008) and its cucumber ortholog *Cs-ACS2*, respectively (Boualem et al. 2008). Both the genes are essential for the development of the female flower and the suppression of male organ development. The *Cm-ACS7* and *Cs-ACS2* transcripts exclusively accumulate in the pistil and ovule, according to in situ mRNA hybridization assays, while its Arabidopsis ortholog, *AT4G26200*, is only expressed in the roots (Yamagami et al. 2003). These results suggest that the acquisition of novel cis regions of the ACS genes may have contributed to the evolution of unisexual flowers in cucurbits. A 454 pyrosequencing analysis of 359,105 expressed sequence tag (EST) sequences from near-isogenic unisexual and bisexual flower buds revealed that six auxin-related genes and three short-chain dehydrogenase or reductase genes are more highly expressed in unisexual flowers (Huang et al. 2009).

7 Transcriptional Control of Bitterness in Cucumber

Regulation of bitterness has been well studied in cucumber. Two transcription factors, responsible for bitterness in fruits and leaves, have been identified by Shang et al. (2014). The bitterness-related loci, which have already been reported, have Bitter (*Bi*) and Bitter fruit (*Bt*) genes that regulated bitterness basically due to

Cucurbitacin C (CuC) in cucumber and related plants. Cucumber genome variation map has been drawn by use of association analysis. An oxidosqualen cyclase gene (*Csa6G088690*) was identified known as cucumber *Bi* gene. To elucidate further, researchers made an ethylmethane sulfonate (EMS)-based mutation-induced library and screened for nonbitter leaves in large number of plants. On sequencing two mutant plants, a single nucleotide polymorphism (SNP) in gene *Csa5G156220* was obtained, which encodes a transcription factor expressed exclusively in leaves. This gene was renamed as Bitter leaf (*Bl*), and that actually regulated the expression of *Bi* gene. A total of 11 variants in homolog of *Bl* (*Csa5G157230*) were found associated with extreme bitterness via local association analysis. The cosegregation with *Bt* gene was observed in F_2 population suggesting that it should be *Bt* gene. It was well established by several workers that the expression of *Bi* gene is regulated by *Bt* gene. A total nine genes have been identified downstream of *Bl* and *Bt* for the biosynthetic pathway for CuC via RNA-interference system.

8 Cucumber Genome-Wide Delineation

Cucumber genome (Huang et al. 2009), melon genome (Garcia-Mas et al. 2012), and multiple other Cucurbitacae genomes (Guo et al. 2020; Kim et al. 2016) have been analyzed and published that elucidated various aspects of genome architecture and functions. The sequence information have opened the doors for various genome-wide expression analysis and identification of gene families and functional characterization at genome scale. As cucumber was the first released genome among cucurbits, all the focus for identification and characterization of gene families was on cucumber. Most of the transcription factors with structural and functional genes were well-described by various studies in cucumber; their transcriptional expression has been instigated under adverse climatic conditions like high salinity, high and low temperatures, hormonal disturbances, etc. in different tissues (roots, leaves, stems, flower buds, and fruits) at different developmental stages. The novel bioinformatic methods are minimizing sequencing errors offering a good-quality genome contiguity on manyfold less costs. The even resequencing of many crops is being done with magnifying exploration of the genome. Simple sequence repeats (SSRs) or microsatellites and other functional markers from cucumber, melon, watermelon, bitter gourd, and spine gourd have been identified (Blanca et al. 2011; Cavagnaro et al. 2010; Zhu et al. 2016; Shukla et al. 2015; Ameen et al. 2022). Other important genomic regions like long intergenic noncoding RNA (lincRNA) have also been identified in cucumber that are at least 200 nucleotide long intergenic transcripts (Hao et al. 2015). Intergenic transcripts encode linc RNAs associated predominantly with the euchromatic or gene-rich regions of the genome with ability to transcription termination and initiation mainly targeting miRNAs. The gene regulation and transcription factor gene activities are together governed by these regulatory repeats. These unique properties compel researchers for identification and characterization of these microsatellites and small RNAs in cucumber and other plant species (Fig. 2).

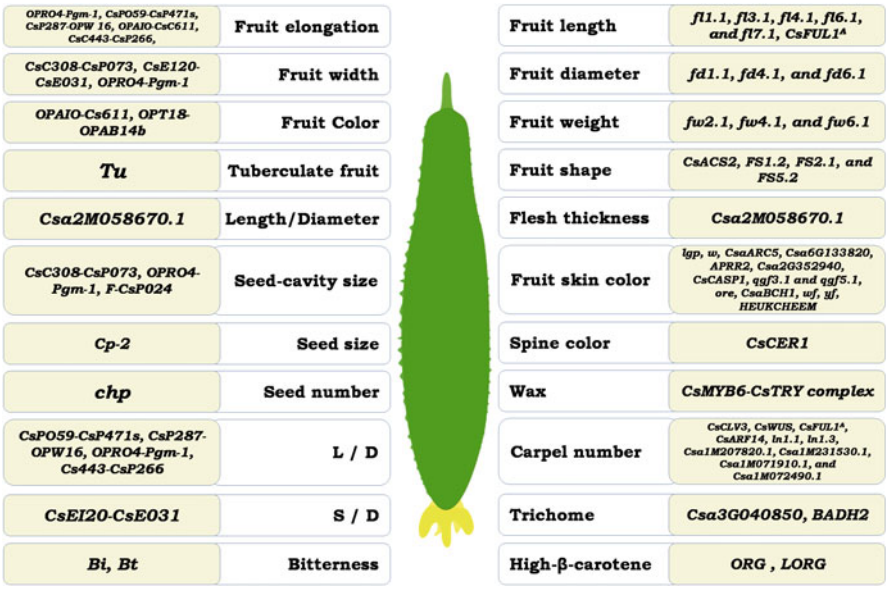


Fig. 2 Some important quality-related traits and responsible genes

9 Transcriptome Analysis in Cucumber

The speedy advancement of high-throughput sequencing tools and genome assembly pipelines has made huge gene annotation information, which offers a prospect for functional studies of genes that are associated to vital traits. RNA-seq technology has made it easy to study various complex traits related to plant growth and development. The technique allows exploring and integrating various gene expression studies and novel gene identification studies. Sex determination, a complex process with many metabolic pathways (e.g., ethylene biosynthesis and other hormone-signaling pathways), is being studied in cucumber since long. These are confirming engagements of strong regulatory processes like ion homeostasis, cell wall synthesis, cytoskeleton modifications, ubiquitination, lipid and sugar metabolism, and gene expression mediated by transcription factors for sex differentiation. Comprehensive analysis of mRNA-Sequence data also enables to identify key genes influenced under various stresses in cucumber. With the aid of high-throughput technologies, it became easy to simultaneously discover and estimate abundance of RNA-Seq data. The novel technology has helped to search new genes as well as to study the transcriptomes. The RNA-Seq can be achieved with the help of algorithms that are not bound to use old annotations and are responsible for alternative transcription and splicing (Trapnell et al. 2010). Prior to this RNA-Seq technology, the experiments were mostly dependent on microarray platform with a limited number of genes plotted on the microarray slides/chips. The major limitation to the

microarray technology is its inability to identify novel transcripts desired for the development of functional molecular marker.

The new technologies like next-generation sequencing (NGS) have ramped the genomics, transcriptomics, and proteomics studies in cucurbits. The availability of cucumber genome has accelerated the molecular studies. The genetic data along with clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9 (CRISPR/Cas9)-based new genome editing techniques are flourishing. Simultaneously the advancement of breeding technologies and sequence-based marker system with genomic and bioinformatics-based tools are also contributing to genetic improvement of cucumber and related cucurbits. This extensive research will lead to better knowledge about cucumber genomics, and scientists will be able to change various traits according to the consumer needs.

The sequence data exploration studies of two cucumber lines PI 308915 (compact vining) and PI 249561 (regular vining) resulted in a total of 200 genes expressed differently, and the SNPs and SSRs among two isogenic lines were also obtained. The protein coding genes were studied by comparison with different types of tissues (leaf, root, stem, female flower, male flower, tendril, ovary, expanded ovary under fertilization, expanded ovary unfertilized, and base part of tendril). The results were compared with previously available protein-coding gene set (Li et al. 2011; Huang et al. 2009). In the study, 8700 genes were found to be structural modifications, where 5300 genes were identified showing similarity with cucumber genome (Li et al. 2011).

As per results of *de novo* transcriptome analysis, SSRs, single nucleotide variants (SNVs) and singly nucleotide molecular markers were useful for vegetable improvement. The construction of genetic map helps to conduct genetic studies of cucumber and related family of plants.

10 Mapping of Fruit Quality-Related Quantitative Trait Loci (QTLs)

The quantitative trait loci (QTLs) control a number of agronomic traits. When one or more sequence-based DNA markers are used in conjunction with an appropriate experimental method and statistical analysis, a quantitative trait locus (QTL) can be described as any region of the genome that is linked to the persistent variation for a trait phenotype in a pertinent reference population. It is determined by analyzing the associated molecular markers for these clusters of sequences governing QTLs for specific traits, helps breeders in genetic studies and Marker Assisted Selection. The ideal populations for QTL mapping are recombinant inbred lines (RILs), near-isogenic lines (NILs), or any segregating wide cross population. RILs and NILs are almost homozygous and have less environmental influences, which increase the accuracy of QTL detection. It is then possible to clone the gene underlying the QTL and use the detected markers in molecular breeding programs.

Similar to a Mendelian hereditary characteristic, the QTLs controlling horticultural traits can be introduced into desirable parents by molecular marker assisted selection

(MAS). Using the F_3 population (GY14 x PI 432860), Wenzel et al. (1995) determined the QTLs for fruit length, fruit diameter, seed cavity size, and color. They noticed that several QTLs overlapped with fruit length and diameter as well as seed cavity size and fruit diameter. It was suggested that QTL overlapping is due to phenotypic positive correlation between fruit diameter and seed cavity size and negative correlation between fruit length and diameter. The QTLs for sex expression, the number of lateral branches, fruit weight, fruit length, and fruit diameter were examined by Serquen et al. (1997). Fruit diameter was discovered to be close to the QTLs for fruit weight, while QTLs underlying lateral branching and fruit diameter were shown to be in the same chromosomal area. Four QTLs (LOD > 3) for sex expression, four for multiple lateral branching, two for earliness, and two for fruit length were also detected. OP-AJ2-F (sex expression), de-OP-L18-2 and BC-403-OP-W7-2 (multiple lateral branching), BC-551-II and II-BC592 (fruit weight), and II-BC592, de-OP-L18-2, and BC551-II were the flanking markers linked to the QTLs (fruit length and diameter). Fazio et al. (2003) developed and studied 171 RILs using the phenotypes having unique alleles for gynoecious (F), determinate (de), standard-sized leaf, the monoecious, indeterminate, and little-leaf (ll) and found additional QTLs for multiple lateral branches, fruit form, and sex expression in addition to confirming the findings of Serquen et al. (1997). In cucurbits, MAS has been successfully demonstrated for phenotypic screening of the *Fom-2* gene (*Fusarium* wilt disease resistance in melon)-associated markers and for multiple lateral branching (*MLB*) in cucumber (Wang et al. 2000; Burger et al. 2003; Fazio et al. 2003). The color of fruits is dependent on concentration and composition of some pigments, mainly flavonoids (especially anthocyanins and chalcones), carotenoids, and chlorophylls. Melon rinds contain a number of colors, e.g., green, orange, white, variegated, and yellow, and are striped too. It is well known that pigments like β-carotene accumulate to contribute orange color (Ma et al. 2022).

11 Metabolic Pathway Studies for Quality Traits in Cucumber

Metabolites are the final products of cell biological regulation process, and metabolomic analysis enables us investigate the relationship between biological processes and plant characteristic. The content of anthocyanins and flavonoids has crucial effect on fruit color and taste. The metabolome data combining with transcriptome profiling discovered genes involved in anthocyanins and flavonols synthesis, thus searching for useful information to illustrate phenomenon of different colors in cucumber fruits (Table 2).

12 Conclusion and Future Perspectives

The diversity in cucumber is mostly found due to simple genetic factors in association with various environmental factors constituting fruit quality, such as bitterness, fruit shape, skin color, spine color, presence or absence of warts and spines, and fruit firmness. Fruit firmness is a crucial factor that is influenced by the

Table 2 Volatile compounds found in cucumber

Functional group	Photochemical
Alkaloids	Hordenine; Piperidine; Isohemiphloin; Betaine; Trigonelline; and Aminophylline
Anthocyanins	Peonidin O-hexoside; Rosinidin O-hexoside; Peonidin; Cyanidin O-syringic acid; Cyanidin O-acetylhexoside; Malvidin 3-O-galactoside; Malvidin 3-O-glucoside (Oenin); Pelargonidin; and Malvidin 3,5-diglucoside (Malvin)
Carbohydrates	D(−)-Threose; Ribulose-5-phosphate; Glucosamine; 2-Deoxyribose 1-phosphate; D(+)-Melezitose O-rhamnoside; Glucarate O-Phosphoric acid; Trehalose 6-phosphate; D-(+)-Sucrose; D(+)-Melezitose; Gluconic acid; D-(+)-Glucono-1,5-lactone; (+)-Glucose; DL-Arabinose; L-Gulonic-γ-lactone; N-Acetyl-D-glucosamine; D-Glucose 6-phosphate; D-Sedoheptuiose 7-phosphate; D-Fructose 6-phosphate; and D-glucoronic acid
Coumarins	N-sinapoyl hydroxycoumarin; O-Feruloyl 4-hydroxylcoumarin; 7-hydroxycoumarin-beta-rhamnoside; Scopoletin (7-Hydroxy-5-methoxycoumarin); 4-Hydroxycoumarin; Esculetin (6,7-dihydroxycoumarin); 6-Methoxy-7,8-DihydroxyCoumarin; Psoralen; and Daphnetin
Flavone	Selgin 5-O-hexoside; Chrysoeriol 5-O-hexoside; Tricin 5-O-acetylglucoside; Selgin O-malonylhexoside; Tricetin O-malonylhexoside; Spinacetin; Syringetin 5-O-hexoside; Luteolin O-sinapoylhexoside; Chrysoeriol O-sinapoylhexoside; Chrysoeriol O-hexosyl-O-rutinoside; Chrysoeriol 7-O-rutinoside; Chrysoeriol O-hexosyl-O-pentoside; Syringetin 7-O-hexoside; Chrysoeriol O-malonylhexoside; Tricin 5-O-hexosyl-O-hexoside; Tricin 7-O-hexosyl-O-hexoside; Tricin O-malonylhexoside; Tricin 7-O-feruloylhexoside; Tricin 7-O-hexoside; Tricin O-sinapoylhexoside; Tricin; Acetyl-eriodictyol O-hexoside; Acacetin O-acetyl hexoside; Chrysoeriol O-hexosyl-O-hexosyl-O-Glucuronic acid; Chrysoeriol O-acetylhexoside; Apigenin 7-O-glucoside (Cosmosiin); Chrysoeriol 7-O-hexoside; Tricin O-sinapic acid; Tricin O-saccharic acid; Tricin 5-O-hexoside; Tricin 5-O-rutinoside; Tricin di-O-hexoside; Tricin O-glucuronic acid; Luteolin; Apigenin 7-O-neohesperidoside (Rhoifolin); Chrysoeriol; Apigenin 7-rutinoside (Isorhoifolin); Baicalein (5,6,7-Trihydroxyflavone); Nobiletin; Tangeretin; Luteolin 7-O-glucoside (Cynaroside); Tricetin; and Butin
Flavanone	Naringenin O-malonylhexoside; Eriodictyol O-malonylhexoside; Hesperetin O-malonylhexoside; Hesperetin 7-O-neohesperidoside (Neohesperidin); Naringenin 7-O-neohesperidoside (Naringin); Naringenin 7-O-glucoside (Prunin); Naringenin; Xanthohumol; Hesperetin 5-O-glucoside; 7-O-Methyleriodictyol; Hesperetin; Hesperetin 7-rutinoside (Hesperidin); Naringenin chalcone; Isosakuranetin-7-neohesperidoside (Poncirin); Afzelechin (3,5,7,4′-Tetrahydroxyflavan); and Homoeriodictyol

(continued)

Table 2 (continued)

Functional group	Photochemical
Flavone C-glycosides	Naringenin C-hexoside; Apigenin C-glucoside; O-methylChrysoeriol 8-C-hexoside; C-hexosyl-chrysoeriol O-hexoside; Apigenin 6-C-hexosyl-8-C-hexosyl-O-hexoside; Hesperetin C-hexosyl-O-hexosyl-O-hexoside; 8-C-hexosyl-hesperetin O-hexoside; Chrysoeriol 6-C-hexoside 8-C-hexoside-O-hexoside; 6-C-hexosyl chrysoeriol O-hexoside; C-hexosyl-apigenin O-feruloylhexoside-O-hexoside; 6-C-hexosyl-hesperetin O-hexoside; 8-C-hexosyl-luteolin O-pentoside; di-C, C-hexosyl-apigenin; 8-C-hexosyl-luteolin O-hexoside; 6-C-hexosyl-apigenin O-sinapoylhexoside; 6-C-hexosyl-apigenin O-feruloylhexoside; 8-C-hexosyl-chrysoeriol O-feruloylhexoside; C-hexosyl-chrysoeriol O-sinapoylhexoside; 8-C-hexosyl-apigenin O-feruloylhexoside; C-hexosyl-apigenin O-p-coumaroylhexoside; 8-C-hexosyl chrysoeriol O-hexoside; Chrysoeriol 8-C-hexoside; 6-C-hexosyl-chrysoeriol O-feruloylhexoside; Eriodictyol C-hexoside; Luteolin C-hexoside; and Isovitexin
Flavonol	MethylQuercetin O-hexoside; Kaempferide; Isorhamnetin O-hexoside; Isorhamnetin 5-O-hexoside; Quercetin 5-O-malonylhexosyl-hexoside; Quercetin 7-O-malonylhexosyl-hexoside; Quercetin 7-O-rutinoside; Isorhamnetin O-acetyl-hexoside; Quercetin O-acetylhexoside; Di-O-methylquercetin; Quercetin 3-O-rutinoside (Rutin); Quercetin; Quercetin 3-alpha-L-arabinofuranoside (Avicularin); Kaempferol 3-O-rutinoside (Nicotiflorin); Myricetin; Isorhamnetin 3-O-neohesperidoside; Isorhamnetin; Kaempferol 3-O-robinobioside (Biorobin); Kaempferol 3-O-glucoside (Astragalin); Dihydromyricetin; Quercetin 4′-O-glucoside (Spiraeoside); Quercetin 3-O-glucoside (Isotrifoliin) Kaempferol 3-O-galactoside (Trifolin); Kaempferol 3-O-rhamnoside (Kaempferin); Fustin; Kaempferol-3-O-robinoside-7-O-rhamnoside (Robinin); Myricetin 3-O-galactoside; and Morin
Flavone C-glycosides	Naringenin C-hexoside; Apigenin C-glucoside; O-methylChrysoeriol 8-C-hexoside; C-hexosyl-chrysoeriol O-hexoside; Apigenin 6-C-hexosyl-8-C-hexosyl-O-hexoside; Hesperetin C-hexosyl-O-hexosyl-O-hexoside; 8-C-hexosyl-hesperetin O-hexoside; Chrysoeriol 6-C-hexoside 8-C-hexoside-O-hexoside; 6-C-hexosyl chrysoeriol O-hexoside; C-hexosyl-apigenin O-feruloylhexoside-O-hexoside; 6-C-hexosyl-hesperetin O-hexoside; 8-C-hexosyl-luteolin O-pentoside; di-C, C-hexosyl-apigenin; 8-C-hexosyl-luteolin O-hexoside; 6-C-hexosyl-apigenin O-sinapoylhexoside; 6-C-hexosyl-apigenin O-feruloylhexoside; 8-C-hexosyl-chrysoeriol O-feruloylhexoside; C-hexosyl-chrysoeriol O-sinapoylhexoside; 8-C-hexosyl-apigenin O-feruloylhexoside; C-hexosyl-apigenin O-p-coumaroylhexoside; 8-C-hexosyl chrysoeriol O-hexoside; Chrysoeriol 8-C-hexoside; 6-C-hexosyl-chrysoeriol O-feruloylhexoside; Eriodictyol C-hexoside; Luteolin C-hexoside; and Isovitexin

(continued)

Table 2 (continued)

Functional group	Photochemical
Lipids_fatty acids	14,15-Dehydrocrepenynic acid; 9-Hydroxy-(10E,12Z,15Z)-octadecatrienoic acid; delta-Tridecalactone; 4-oxo-9Z,11Z,13E,15E-octadecatetraenoic acid; Punicic acid; Octadecadien-6-ynoic acid; Octadeca-11E,13E,15Z-trienoic acid; 4-Hydroxysphinganine; 8,15-DiHETE; Lauric acid (C12:0); Myristoleic acid (C14:1); 9,10-EODE; 13-HOTrE; 9-HOTrE; 9-KODE; 13-HpOTrE(r); 9-HpOTrE; 13-HOTrE(r); 12,13-EODE; 13-HPODE; and α-Linolenic acid
Lipids_glycerolipids	MAG (18:4) isomer1; DGMG (18:2) isomer2; MAG (18:3) isomer5; DGMG (18:2) isomer1; DGMG (18:2) isomer3; MAG (18:2) isomer1; MAG (18:4) isomer2; MAG (18:1) isomer2; AG (18:2); MAG (18:4) isomer3; MAG (18:3) isomer3; MGMG (18:2) isomer1; MAG (18:3) isomer4; DGMG (18:1); MAG (18:3) isomer2; MAG (18:1) isomer1; MGMG (18:2) isomer2; and MAG (18:3) isomer1
Lipids_glycerophospholipids	LysoPC 16:2; LysoPC 16:1; PC 16:1/14:1; LysoPC 16:1 (2n isomer); LysoPC 18:2; LysoPC 18:3; LysoPC 16:0; LysoPE 18:1 (2n isomer); LysoPC 18:1 (2n isomer); LysoPC 16:2 (2n isomer); LysoPE 14:0; LysoPC 18:3 (2n isomer); LysoPC 14:0; LysoPC 18:2 (2n isomer); LysoPE 18:2 (2n isomer); LysoPC 18:1; LysoPE 18:0; LysoPC 19:0; LysoPC 15:1; LysoPC 15:0; LysoPC 18:0 (2n isomer); LysoPC 17:0; LysoPE 16:0; LysoPE 18:1; LysoPC 20:4; LysoPC 14:0 (2n isomer); LysoPC 16:0 (2n isomer); LysoPC 18:0; LysoPC 20:1 (2n isomer); LysoPC 20:1; LysoPE 14:0 (2n isomer); and LysoPE 16:0 (2n isomer)
Phytohormones	Kinetin 9-riboside; trans-zeatin N-glucoside; trans-zeatin 9-O-glucoside; Salicylic acid O-glucoside; Methyl jasmonate; Indole 3-acetic acid (IAA); (+)-Jasmonic acid (JA); Salicylic acid (SA); N6-Isopentenyladenine (iP); trans-Zeatin (tZ); cis-Zeatin (cZ); Gibberellin A1 (GA1); Gibberellin A4 (GA4); (+)-cis,trans-Abscisic acid (ABA); N-[(−)-Jasmonoyl]-(L)-Isoleucine (JA-L-Ile); GIBBERELLIN A15; Gibberellin A20; and GIBBERELLIN A9
Nucleotide and its derivates	N2-methylguanosine; Xanthine; Adenosine 3′-monophosphate; Nicotinic acid adenine dinucleotide; Inosine 5′-monophosphate; iP7G; Adenosine 5′-monophosphate; Guanosine 5′-monophosphate; Cyclic AMP; Uridine 5′-diphospho-D-glucose; 2′-Deoxyinosine-5′-monophosphate; 6-Methylmercaptopurine; Adenosine O-ribose; Succinyladenosine; Adenosine; Thymine; Hypoxanthine; Cytosine; Adenine; 5-Methylcytosine; β-Nicotinamide mononucleotide; 2-Hydroxy-6-aminopurine; Uracil; Thymidine; Uridine; Guanine; Inosine; Guanosine; Deoxyguanosine; Deoxycytidine; Xanthosine; 8-Hydroxyguanosine; 2′-Deoxycytidine-5′-monophosphate; 1-Methyladenosine; 5′-Deoxy-5′-(methylthio)adenosine; Guanosine monophosphate; Flavin adenine dinucleotide (FAD); 7-Methylxanthine; Uridine 5′-diphosphate; 1-Methyladenine; Deoxyribose 5-phosphate; Cytidine 5′-monophosphate

(continued)

Table 2 (continued)

Functional group	Photochemical
	(Cytidylic acid); 2′-Deoxyadenosine-5′-monophosphate; Uridine 5′-monophosphate; 1-methylguanidine; N6-Succinyl Adenosine; Cytidine; Deoxyadenosine; 2-(dimethylamino)guanosine; β-Pseudouridine; H; and ypoxanthine-9-β-D-arabinofuranoside
Organic acids	Sinapoyl malate; Xanthurenic acid; Methylglutaric acid; Phosphoric acid; 4-Hydroxy-3-methoxymandelate; Argininosuccinate; Citramalate; Diethyl phosphate; 2-Isopropylmalate; Rosmarinic acid; 3-Hydroxybutyrate; Kynurenic acid; Ethyl 3,4-Dihydroxybenzoate (Ethyl protocatechuate); lutaric acid; Adipic acid; Azelaic acid; Azelaic Acid; Sebacate; 2-Methylsuccinic acid; Maleic acid; 6-Aminocaproic acid; Terephthalic acid; Phthalic acid; 4-Guanidinobutyric acid; 2-Hydroxyisocaproic acid; Dl-2-Aminooctanoic acid; 4-Acetamidobutyric acid; Shikimic acid; Methylmalonic acid; 2-Picolinic acid; D-Pantothenic dcid; 3-Hydroxy-3-methyl butyric acid; D-Erythronolactone; Succinic acid; Creatine; Suberic acid; L-(+)-Tartaric acid; L(−)-Malic acid; Citric acid; (S)-(−)-2-Hydroxyisocaproic acid; Fumaric acid; Citraconic acid; 3-Aminosalicylic acid; Dodecanedioic aicd; A-Ketoglutaric acid; α-Hydroxyisobutyric acid; Cis-Aconitic acid; Oxoadipic acid; 3-Hydroxypropanoic acid; Guanidinoethyl sulfonate; 4-Hydroxybenzoic acid; trans-Citridic acid; γ-aminobutyric acid; ethylmalonate; Taurocholic acid; 2-(Formylamino)benzoic acid; Aminomalonic acid; (Rs)-Mevalonic acid; trans,trans-Muconic acid; 2-Methylglutaric acid; 5-hydroxyhexanoic acid; and D-Xylonic acid
Terpenoids	Phytocassane D; Cucurbitacin E; Cucurbitacin I; Nomilin; and Limonin
Vitamins	Pyridoxine O-glucoside; Niacinamide; 4-Pyridoxic acid O-hexoside; D-Pantothenic acid; Thiamine; Menaquinone (K2); Pyridoxal 5′-phosphate; Pyridoxine 5′-phosphate; Pyridoxine; Nicotinic Acid Methyl Ester (Methyl Nicotinate); Riboflavin; L-ascorbate; Orotic acid; Biotin; Vitamin D3; 4-Oxoretinol; All-trans-13,14-dihydroretinol; and Pantetheine

[a]Table adopted from Wang et al. (2020)

firmness of the flesh and the seed cavity (endocarp). Various genetic and metabolic datasets for cucumbers are now accessible for cucurbit breeding and biological study. Cucumber serves as a model plant for the cucurbit family because of the strong collinearity between it and the majority of other cucurbits. The cucumber's genome has just undergone an upgrade, making it possible to use this knowledge for focused breeding efforts. The metabolic profile identified the metabolic intermediates as well as potential dietary and nondietary components. Alkaloids, tannins, flavonoids, steroids, phlobatannins, and saponins are just a few of the physiologically active, nonnutritive phytochemicals that are present in both cucumber fruits and seeds at the same time. Knowledge of the mechanisms

underlying sex determination can speed up the process of obtaining desired fruit quality and will help breeders meet other objectives like producing cucumbers with a high and steady yield.

References

Abiodun OA, Adeleke RO (2010) Comparative studies on nutritional composition of four melon seeds varieties. Pak J Nutr 9. https://doi.org/10.3923/pjn.2010.905.908

Agarwal M, Kumar A, Gupta R, Upadhyaya S (2012) Extraction of polyphenol, flavonoid from *Emblica officinalis*, *Citrus limon*, *Cucumis sativus* and evaluation of their antioxidant activity. Orient J Chem 28(2):993

Ameen G, Prakash V, Ram N, Sandilya VK, Kumar A, Tiwari JK (2022) Standardization of protocol for genomic DNA extraction and microsatellite marker (SSR, ISSR) analysis in spine gourd (*Momordica dioica*). Curr Hortic 10(2):38–40

Arya V, Thakur N, Kashyap C (2012) Ethnomedicinal plants of Kangra Valley of Himachal Pradesh. India J Trop Med Plants 13(2)

Begum HA, Asad F, Sadiq A, Mulk S, Ali K (2018) Antioxidant, antimicrobial activity and phytochemical analysis of the seeds extract of *Cucumis sativus* Linn. Pure Appl Biol 8(1):433–441

Blanca JM, Cañizares J, Ziarsolo P, Esteras C, Mir G, Nuez F, Garcia-Mas J, Picó MB (2011) Melon transcriptome characterization: simple sequence repeats and single nucleotide polymorphisms discovery for high throughput genotyping across the species. Plant Genome 4(2)

Boualem A, Fergany M, Fernandez R, Troadec C, Martin A, Morin H, Sari M-A, Collin F, Flowers JM, Pitrat M (2008) A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. Science 321(5890):836–838

Burger Y, Katzir N, Tzuri G, Portnoy V, Saar U, Shriber S, Perl-Treves R, Cohen R (2003) Variation in the response of melon genotypes to *Fusarium oxysporum* f. sp. *melonis* race 1 determined by inoculation tests and molecular markers. Plant Pathol 52(2):204–211

Campbell JM (2008) Pharmacy in ancient Egypt. Egyptian mummies and modern science. Cambridge University Press, United Kingdom, pp 216–233

Cavagnaro PF, Senalik DA, Yang L, Simon PW, Harkins TT, Kodira CD, Huang S, Weng Y (2010) Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). BMC Genomics 11(1):1–18

Chen JF, Zhang S, Zhang XG (1994) The xishuangbanna gourd (*C. sativus* var. *xishuangbannesis* Qi et Yuan), a traditionally cultivated plant of the Hanai people, Xishuangbanna, Yunnan, China. Cucurbit Genet Coop Rpt 17

Chen JF, Isshiki S, Tashiro Y, Miyazaki S (1995) Studies on a wild cucumber from China (*Cucumis hystrix* Chakr.). I. Genetic distances between C. hystrix and two cultivated cucumis species (*C. sativus* L. and *C. melo* L.) based on isozyme analysis. J Jpn Soc Hort Sci 64

Chiej R (1984) Encyclopaedia of medicinal plants. MacDonald and Company Limited Maxwell House, London, pp 128–131

Dane F (1991) Cytogenetics in genus *Cucumis*. In: Tsuchiya T, Gupta PK (eds) Chromosome engineering in plants: genetics, breeding, and evolution, vol 1. Elsevier, Amsterdam, pp 201–214

FAOSTAT (2018) http://www.fao.org/faostat/en/#data. Accessed 12 Mar 2018

Fazio G, Chung SM, Staub JE (2003) Comparative analysis of response to phenotypic and marker-assisted selection for multiple lateral branching in cucumber (*Cucumis sativus* L.). Theor Appl Genet 107(5):875–883

Field B, Osbourn AE (2008) Metabolic diversification – independent assembly of operon-like gene clusters in different plants. Science 320(5875):543–547

Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, González VM, Hénaff E, Câmara F, Cozzuto L, Lowy E (2012) The genome of melon (*Cucumis melo* L.). Proc Natl Acad Sci U S A 109(29):11872–11877

Gopalakrishnan SB, Kalaiarasi T, Vadivel E (2014) In vitro anthelmintic screening comparison of various crude extracts of the fruits of *Cucumis trigonus* Roxb. and *Cucumis sativus* Linn. World J Pharmaceut Res 3(4):582–590

Grieve M (1998) A modern herbal—the medicinal, culinary, cosmetic and economic properties, cultivation and folklore of herbs, grasses, fungi, shrubs and trees with all their modern scientific uses. Tiger Books International, London

Guo J, Xu W, Hu Y, Huang J, Zhao Y, Zhang L, Huang C-H, Ma H (2020) Phylotranscriptomics in Cucurbitaceae reveal multiple whole-genome duplications and key morphological and molecular innovations. Mol Plant 13(8):1117–1133

Hao Z, Fan C, Cheng T, Su Y, Wei Q, Li G (2015) Genome-wide identification, characterization and evolutionary analysis of long intergenic noncoding RNAs in cucumber. PLoS One 10(3): e0121800

Huang S, Li R, Zhang Z (2009) The genome of the cucumber, *Cucumis sativus* L. Nat Genet 41. https://doi.org/10.1038/ng.475

Jeffrey C (1980) A review of the Cucurbitaceae. Botanical Journal of the Linnean Society 81 (3):233–247. https://doi.org/10.1111/j.1095-8339.1980.tb01676.x

Kim HA, Shin A-Y, Lee M-S, Lee H-J, Lee H-R, Ahn J, Nahm S, Jo S-H, Park JM, Kwon S-Y (2016) De novo transcriptome analysis of *Cucumis melo* L. var. makuwa. Mol Cells 39(2):141

Kirkbride JH (1993) Biosystematic monograph of the genus *Cucumis* (Cucurbitaceae). Parkway Publishers, Boone, pp 1–159

Köller W, Frevert J, Kindl H (1979) Albumins, glyoxysomal enzymes and globulins in dry seeds of *Cucumis sativus*. Qual Quant Anal 360(1):167–176

Kumar D, Kumar S, Singh J, Vashistha B, Singh N (2010) Free radical scavenging and analgesic activities of *Cucumis sativus* L. fruit extract. J Young Pharmacists 2(4):365–368

Li Y, Yang L, Pathak M, Li D, He X, Weng Y (2011) Fine genetic mapping of cp: a recessive gene for compact (dwarf) plant architecture in cucumber, *Cucumis sativus* L. Theor Appl Genet 123

Lim T (2012) *Cucumis sativus* edible medicinal and non-medicinal plants. Springer, Dordrecht, pp 239–249

Ma L, Wang Q, Zheng Y, Guo J, Yuan S, Fu A, Bai C, Zhao X, Zheng S, Wen C (2022) Cucurbitaceae genome evolution, gene function, and molecular breeding. Hort Res 9. https://doi.org/10.1093/hr/uhab057

Mallik J, Das P, Das S (2013) Pharmacological activity of *Cucumis sativus* L. – a complete overview. Asian J Pharmaceut Res Dev 1(1):1–6

Pandey S, Kujur SN (2022) Importance, distribution, botany and genetics. In: The Cucumber genome. Springer, Cham, pp 1–13

Pandey S, Ansari WA, Pandey M, Singh B (2018) Genetic diversity of cucumber estimated by morpho-physiological and EST-SSR markers. Physiol Mol Biol Plants 24(1):135–146

Patil MV, Kandhare A, Bhise S (2011) Pharmacological evaluation of ameliorative effect of aqueous extract of *Cucumis sativus* L. fruit formulation on wound healing in Wistar rats. Chron Young Sci 2(4):207–207

Rudich J, Halevy A, Kedar N (1972) Ethylene evolution from cucumber plants as related to sex expression. Plant Physiol 49(6):998–999

Sahu T, Sahu J (2015) *Cucumis sativus* (cucumber): A review on its pharmacological activity. J Appl Pharmaceut Res 3(1):04–09

Serquen FC, Bacher J, Staub J (1997) Mapping and QTL analysis of horticultural traits in a narrow cross in cucumber (*Cucumis sativus* L.) using random-amplified polymorphic DNA markers. Mol Breed 3(4):257–268

Shang Y, Ma Y, Zhou Y, Zhang H, Duan L, Chen H, Zeng J, Zhou Q, Wang S, Gu W (2014) Biosynthesis, regulation, and domestication of bitterness in cucumber. Science 346(6213): 1084–1088

Shibuya M, Adachi S, Ebizuka Y (2004) Cucurbitadienol synthase, the first committed enzyme for cucurbitacin biosynthesis, is a distinct enzyme from cycloartenol synthase for phytosterol biosynthesis. Tetrahedron 60(33):6995–7003

Shukla A, Singh V, Bharadwaj D, Kumar R, Rai A, Rai A, Mugasimangalam R, Parameswaran S, Singh M, Naik P (2015) De novo assembly of bitter gourd transcriptomes: gene expression and sequence variations in gynoecious and monoecious lines. PLoS One 10(6):e0128331

Sotiroudis G, Melliou E, Sotiroudis TG, Chinou I (2010) Chemical analysis, antioxidant and antimicrobial activity of three Greek cucumber (*Cucumis sativus*) cultivars. J Food Biochem 34:61–78

Staub JE, Robbins MD, Wehner TC (2008) Cucumber. In: Prohens J, Nuez F (eds) Vegetables I: Asteraceae, Brassicaceae, Chenopodiaceae, and Cucurbitaceae. Springer, New York

Tang J, Meng X, Liu H, Zhao J, Zhou L, Qiu M, Zhang X, Yu Z, Yang F (2010) Antimicrobial activity of sphingolipids isolated from the stems of cucumber (Cucumis sativus L.). Molecules 15:9288–9297. https://doi.org/10.3390/molecules15129288

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28(5):511–515

Venkatachalam P, Sangeetha P, Geetha N, Sahi SV (2015) Phytofabrication of bioactive molecules encapsulated metallic silver nanoparticles from *Cucumis sativus* L. and its enhanced wound healing potential in rat model. J Nanomater 16(1):241

Wang HZ, Zhao PJ, Zhou XY (2000) Regeneration of transgenic cucumber plants with a WMV-2 CP gene. Acta Phytophysiol Sin 26(3):267–272

Wang M, Chen L, Liang Z, He X, Liu W, Jiang B, Yan J, Sun P, Cao Z, Peng Q (2020) Metabolome and transcriptome analyses reveal chlorophyll and anthocyanin metabolism pathway associated with cucumber fruit skin color. BMC Plant Biol 20(1):1–13

Wenzel G, Kennard WC, Havey MJ (1995) Quantitative trait analysis of fruit quality in cucumber: QTL detection, confirmation, and comparison with mating-design variation. Theoretical and Applied Genetics 91:53–61. https://doi.org/10.1007/BF00220858

Whitaker TW, Davis GN (1962) Cucurbits. Botany, cultivation, and utilization. Cucurbits. Botany, cultivation, and utilization. Interscience Publishers, New York, pp 1–250

Yamagami T, Tsuchisaka A, Yamada K, Haddon WF, Harden LA, Theologis A (2003) Biochemical diversity among the 1-amino-cyclopropane-1-carboxylate synthase isozymes encoded by the Arabidopsis gene family. J Biol Chem 278(49):49102–49112

Zhao L, Huang Y, Zhou H, Adeleye AS, Wang H, Ortiz C, Mazer SJ, Keller AA (2016) GC-TOF-MS based metabolomics and ICP-MS based metallomics of cucumber (*Cucumis sativus*) fruits reveal alteration of metabolites profile and biological pathway disruption induced by nano copper. Environ Sci Nano 3(5):1114–1123

Zhu W-Y, Huang L, Chen L, Yang J-T, Wu J-N, Qu M-L, Yao D-Q, Guo C-L, Lian H-L, He H-L (2016) A high-density genetic linkage map for cucumber (*Cucumis sativus* L.): based on specific length amplified fragment (SLAF) sequencing and QTL analysis of fruit traits in cucumber. Front Plant Sci 7:437

Sweetpotato: Nutritional Constituents and Genetic Composition

Guilherme Silva Pereira, Victor Acheampong Amankwaah, Mercy Ketavi, Bonny Michael Oloka, Aswathy G. H. Nair, Ana Paula da Mata, Carla Cristina da Silva, Iara Gonçalves dos Santos, João Ricardo Bachega Feijó Rosa, and Hugo Campos

Contents

G. S. Pereira (✉) · A. P. da Mata · C. C. da Silva · I. G. dos Santos · J. R. B. F. Rosa
Department of Agronomy, Federal University of Viçosa, Viçosa, Brazil
e-mail: g.pereira@ufv.br; ana.mata@ufv.br; carla.silva1@ufv.br; iara.santos@ufv.br; joao.r.rosa@ufv.br

V. A. Amankwaah
Biotechnology, Seed and Post-Harvest Division, Crops Research Institute, Kumasi, Ghana
e-mail: v.amankwaah@cropsresearch.org

M. Ketavi
Research Technology Support Facility, Michigan State University, East Lansing, MI, USA
e-mail: kitavime@msu.edu

B. M. Oloka
Department of Horticultural Science, North Carolina State University, Raleigh, NC, USA
e-mail: boloka@ncsu.edu

A. G. H. Nair
Division of Crop Improvement, Central Tuber Crops Research Institute, Thiruvananthapuram, India
e-mail: aswathyghnair07@gmail.com

H. Campos
International Potato Center, Lima, Peru
e-mail: h.campos@cgiar.org

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_58

Abstract

Sweetpotato, *Ipomoea batatas* ($2n = 6x = 90$), is a polyploid, outcrossing species considered a staple food in many developing countries due to its starch-rich storage roots. In addition to energy, roots or leaves can provide minerals (such as Fe and Zn), and vitamins (such as A and C). β-carotene- or anthocyanin-rich varieties and their respective orange and purple fleshed roots have attracted great attention given the increased nutraceutical properties they present. Originally from Latina America, several studies have been conducted to assess germplasm diversity around the world using both morphological and molecular descriptors. Similarly, there have been some attempts to describe genetic architecture of traits of interest such as dry matter content and resistance to sweetpotato virus disease and nematodes. However, only recently, major developments in molecular tools have allowed the genetic mapping of major quantitative trait loci and genome-wide-based characterization of population structure. Releasing new varieties that show high yield under challenging environments combined with resistance to pests and diseases plus nutritious characteristics is a long-lasting process. Recent advances such as the availability of diploid, wild relative genomes, and their associated bioinformatics tools have facilitated functional studies. Likewise, the ability of detecting single nucleotide polymorphisms and measuring the allele

dosage have facilitated genetic studies in sweetpotato and have paved the way to the implementation of genomics-assisted breeding.

Keywords

Superfood · Vitamins · Minerals · Genetics · Diversity · Breeding

1 Introduction

Sweetpotato [*Ipomoea batatas* (L.) Lam.] belongs to the Convolvulaceae or morning glory family of the genus *Ipomoea*, where it is the only economically important species. It is an autohexaploid with 90 somatic chromosomes ($2n = 6x = 90$), with a large genome size of about ~4.4 Gb (Yang et al. 2017), and complex inheritance pattern. According to Food and Agriculture Organization of the United Nations, its estimated global production of 90 million metric tons in 2020 is led by Asia and sub-Saharan Africa (SSA), accounting for most of the global sweetpotato production with 83% production in Asia and 14.6% in Africa. In SSA, the crop is a major starch staple and a source of calories, fiber, vitamins, and mineral requirements for humans and domestic animals (Low et al. 2017). In fact, the various parts of sweetpotato crop have been reported to contain phenolics, dietary fiber, and mineral nutrients. Sweetpotato leaves have been reported to contain more polyphenols compounds and at least 15 anthocyanins compared to other commercial vegetables such as spinach, cabbage, and lettuce (Nguyen et al. 2021). Its roots are enriched with secondary metabolites of immense nutritional value and high sensory versatility in terms of taste, texture, and flesh color (white, cream, yellow, orange, and purple).

Orange-fleshed sweetpotato (OFSP) is a very nutritious food, as it is a very good source of β-carotene, iron, potassium, vitamin C, fiber, and protein (Low et al. 2017). Sweet OFSP clones with low dry matter (<20%) are the preferred types in the USA and Europe, whereas the less sweet, white or cream fleshed clones are the preferred types in much of SSA. Due to their reduced vitamin A content, these types are not as nutritious as the OFSP types. Therefore, the breeding product profile of most SSA sweetpotato breeding programs are tailored toward development of higher dry matter and less sweet OFSP, to address the ever-growing challenge of vitamin A deficiency needs of the population to prevent malnutrition and enhance nutrition and food security. The purple-fleshed sweetpotato (PFSP) varieties are anthocyanin rich with an attractive color with more than 20 known anthocyanins in different varieties (Wang et al. 2018) which are the specialty type in Asia.

Carotenoid biosynthesis and degradation and/or plastid sink strength processes occur naturally in sweetpotato and can be readily selected in breeding populations. Nonetheless, the genetic improvement of sweetpotato for consumer preferred quality traits such as storage root dry matter, starch, sugar, and β-carotene contents has been

difficult. This could be attributed to the complex polyploid genome of sweetpotato, negatively correlated traits, and its outcrossing nature resulting in very challenging breeding (Cervantes-Flores et al. 2011; Truong et al. 2018). There is always a constant need to explore different mechanisms that are most often governed by many genes. In the case of OFSP, the negative starch/β-carotene correlation and the yet undefined textural characteristics have limited the actual adoption of improved OFSP varieties. Therefore, comprehensive analysis of the genetic architecture of the negative association between starch and β-carotene could aid in advances of breaking this linkage as an important objective of breeding programs targeting sweetpotato for food and nutritional security.

2 Description of Nutritional Constituents

2.1 Detailed Chemical Composition

Sweetpotato storage roots contain carbohydrates, proteins, minerals, and vitamins. The major constituents of sweetpotato storage roots are carbohydrates in which the principal component is starch (Woolfe 1992). In addition to starch, the storage roots of sweetpotato provide free sugars such as fructose, glucose, and sucrose in raw and cooked as well as maltose that is usually contained in cooked storage roots (Kitahara et al. 2017; Truong et al. 2018). Other carbohydrates include pectin, hemicellulose, and cellulose, which have been reported to affect the textural properties of storage roots (Woolfe 1992). The vitamin component consists of vitamins C, K, E, and several B vitamins (Truong et al. 2018). The OFSP varieties contain β-carotene which is converted into vitamin A in the body. Research has shown that the darker the orange color in the root flesh, the more β-carotene is present (Low et al. 2017). Sweetpotato root mineral composition includes potassium, phosphorus, and other minerals. Detailed mineral and vitamin composition per 100 g of fresh weight can be found in Table 1.

Storage roots of sweetpotato vary widely in nutrient composition depending on cultivar, growing conditions, maturity, and storage (Truong et al. 2018). Also, different varieties of sweetpotato show variations of chemical composition influenced by dry matter content and storage root flesh color. OFSP varieties, for example, usually have low dry matter content (18–25%), high β-carotene level, and sweetness due to relatively high sugar content. OFSP varieties with low dry matter content (~20%) are the major types of varieties produced in the USA (Cervantes-Flores et al. 2011). These varieties are usually sweet with high sugar (15–20% relative to dry matter – DM) and low starch (45–55% DM) contents. A number of studies have reported negative correlations between starch and dry matter content. In most parts of SSA, the preferred types are cream- or white-fleshed varieties with high dry matter content (28–30%) that have relatively low sweetness (Cervantes-Flores et al. 2011), low sugar (10–20% DM), and high starch (50–80% DM) contents. PFSP varieties, on the other hand, contain high anthocyanin content (Truong et al. 2018). Variability in sugar content is influenced by the conversion of starch into sugars during curing and postharvest storage (Truong et al. 2018). The storage roots of

Table 1 Mineral and vitamin composition of sweetpotato storage roots (per 100 g of fresh weight[a])

Component	Orange-fleshed	Yellow-fleshed	White/cream-fleshed	Purple-fleshed
Minerals (mg)				
Calcium	24.4–100	23.0–25.0	24.0–39.0	18.5–23.4
Iron	0.63–1.26	0.53–1.12	0.1–1.30	1.28–1.30
Magnesium	3.0–37.0	22.5–23.1	15.0–31.0	NA[b]
Potassium	115–334	315–327	191–324	217–438
Phosphorus	12.5–51.0	44.7–46.3	28.0–47.0	NA
Sodium	25.0–31.0	29.0–34.3	31.8–34.0	NA
Zinc	0.23–0.93	0.25–0.27	0.16–0.60	0.23–0.80
Vitamins (mg)				
Ascorbic acid (C)	9.72–24.69	5.82–36.18	4.1–29.86	4.86–30.25
Thiamin (B1)	0.06–0.08	0.06–0.31	0.09–0.13	0.09
Riboflavin (B2)	0.02–0.08	0.01–0.02	0.01–0.02	0.01
Niacin (B3)	0.43–0.78	0.94–1.91	0.54–1.09	0.57
Pantothenic acid (B5)	0.59–1.25	0.43–0.58	0.62–0.84	0.74
Pyridoxine (B6)	0.18–0.25	0.14–0.16	0.08–0.22	0.085
β-carotene[c]	1.66–31.44	0.67–7.47	0–0.55	0–5.66

[a]Ranges reported to the nearest decimal, based on values reported (Neela and Fanta 2019)
[a]Not available
[c]Provitamin A

sweetpotato contain reasonable amounts of protein (~5% DM). In addition, the storage roots also contain α- and β-amylases with β-amylase being the most abundant (Amankwaah 2019).

Variations in chemical composition have also been reported to exist in raw storage roots and cooked ones. Heating sweetpotato storage roots to a temperature of 75 °C retains a relatively high percentage of endogenous amylases required for starch conversion and free sugar formation in the cooked product. It has been observed that increasing temperature and heating time generally reduces β-amylase activity and enhances maltose formation, thereby increasing sweetness (Amankwaah 2019). Major changes were observed in starch content composition in raw storage roots after baking, which demonstrates the fact that starch is hydrolyzed in raw sweetpotatoes to sugars. The major product obtained after starch hydrolysis in cooked products is maltose. Amankwaah (2019) observed that production of maltose caused a big difference in total sugars between raw and baked treatments.

2.2 Chemical Type, Structure, and Biochemical Pathways of Production

The starch contained in storage roots of sweetpotato consists of linear or slightly branched amylose and highly branched amylopectin. The amylose content of sweetpotato starch is approximately 20–30% while its amylopectin is 70–80%. The

size of starch granules is variety-based and determined by starch content, chemical properties, and amylopectin/amylase ratio as well as supramolecular structures (Noda et al. 2009). Starch granule distribution is an important characteristic and vital factor for the quality of final products derived from sweetpotato. Several factors affect the formation of different granules including growth and agronomic management as well as gene expression mechanisms, especially those associated with starch biosynthesis. In sweetpotato, starch biosynthesis is a complex and highly regulated process that requires coordinated activities of multiple enzymes including AGPase, starch synthase (SS), starch branching enzyme (SBE), and starch debranching enzyme (DBE). The structure and properties of starch are controlled by starch biosynthetic enzymes and regulators, such as granule-bound starch synthase I (GBSSI) is the key enzyme in the biosynthesis of amylose. On the other hand, soluble SS, SBEs I and II, and DBE act in concert to produce amylopectin (Kitahara et al. 2017).

Carotenoid pigments are widespread in nature. These molecules in plants play significant roles in light harvesting of photosynthetic reaction centers. Furthermore, in conjunction with chlorophyll in the chloroplast, cells are protected from excessive light by absorbing blue-green wavelengths (Kang et al. 2017). Different carotenoid types are contained in the storage roots of sweetpotato including β-carotene, β-crypthoxanthin, zeaxanthin, violaxan thin, and other unknown carotenoids. However, β-carotene is a major component in yellow- and orange-fleshed sweetpotato (Kang et al. 2017).

The carotenoids metabolic pathway and the function of the biosynthetic enzymes involved have been well-studied in higher plants. They are synthesized by the 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylose 5-phosphate pathway (MEP/DOXP pathway) of isoprenoid biosynthesis. From isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), two molecules of geranylgeranyl diphosphate (GGPP) are produced by geranylgeranyl pyrophosphate synthase (GGPS). Through the condensation of two GGPP molecules and phytoene, the first C40 carotenoid is formed, a reaction catalyzed by phytoene synthase (PSY). Phytoene is then processed by a series of enzymes including phytoene desaturase (PDS) and ξ-carotene desaturase (ZDS) to yield lycopene. Lycopene is further processed by two carotenoid biosynthetic pathways, the α-branch pathway (from α-carotene to lutein) and the β-branch pathway, and both play distinct roles in photo-protection (Kang et al. 2017). Subsequently, α-carotene and β-carotene are modified by hydroxylation, epoxidation, or isomerization to express a variety of structural features. Lutein is then synthesized in the α-branch pathway by α-carotene ε-ling hydroxylase (CHY-ε). On the other hand, in the β-branch pathway, hydroxylation by β-hydroxylase (CHY-β) converts β-carotene to zeaxanthin, and then zeaxanthin epoxidase (ZEP) mediates the formation of violaxanthin. Violaxanthin is converted to neoxanthin by neoxanthin synthase (NXS) (Kang et al. 2017).

2.3 Medicinal Properties and Functions in Relation to Human Health

A wide variety of roots and tubers plays a major role in human diet, animal feed, and industrial raw materials. Sweetpotato is now being promoted as a “superfood” for

good health globally since it is a rich source of carbohydrates, minerals, and vitamins (Low et al. 2017). Sweetpotato is recognized as a secondary staple food, and it is known to play significant role in the diet of humans in many underdeveloped countries (Neela and Fanta 2019). Sweetpotato has been reported to be different from other staple food crops as it has balanced nutritional composition in addition to adaptability in wider topography, good productivity in short durations, and ability to grow on marginal soils. Awareness of the high nutritional value of sweetpotato is driving increasing consumer demand for the crop among health-conscious consumers globally (Grüneberg et al. 2015). In many developed countries, sweetpotato has been reevaluated as a health-promoting food, because of its high content of nutrients in balanced and functional components, such as anthocyanins, carotenoids, and phenolic compounds, which have antioxidant properties.

In addition to nutrients, sweetpotato roots are regarded as functional food and provide physiological benefits. They contain phytochemical compounds that include carotenoids, tocopherols, phenolic compounds, tannins, flavonoids, saponins, and anthocyanins, which differ among varieties based on the storage root flesh color (Woolfe 1992). Either single or synergistically, these bioactive phytochemicals provide the human body with antioxidant, cardioprotective, antidiabetic, hepatoprotective, neuroprotective, anti-inflammatory, antimicrobial, and bowel-regulation properties (Panda and Sonkamble 2012). Storage roots, when consumed containing these bioactive chemicals, result in boosting of the immune systems to fight diseases and subsequent promotion of a healthy life and longevity. Furthermore, these bioactive phytochemicals serve as potential sources of antioxidants that can scavenge free radicals and reduce or inhibit cellular damages. Additionally, reduction in oxidation stress results in disease prevention and better health (Anbuselvi and Muthumani 2014).

Significant improvement in human health can be achieved when biofortified crops such as sweetpotato are consumed regularly. In fact, OFSP varieties are rich in provitamin A carotenoids, mainly in the form of β-carotene (Neela and Fanta 2019). They have potential to prevent vitamin A deficiency and are good for consumption by children and pregnant women. Research on nutraceutical properties of PFSP varieties indicated that the extracted anthocyanins showed strong radical scavenging activity, antimutagenic activity, and significant reduction in high blood pressure and liver injury in rats (Truong et al. 2018).

2.4 Cultural Methods of Nutraceutical Improvement

The process of increasing vitamins and minerals in a crop attributed to plant breeding, transgenic techniques, or agronomic practices is termed biofortification. It is a feasible and cost-effective mean of delivering micronutrients to populations that may have limited access to diverse diets and other micronutrient interventions (Bouis and Saltzman 2017). Increasing nutrition levels in staple crops to target levels, required for improving human nutrition without compromising yield or farmer-preferred agronomic traits, could be achieved through conventional plant breeding. It involves screening germplasm for available genetic diversity, prebreeding parental genotypes, developing and testing micronutrient dense germplasm, selecting, and advancing

generations for product development. Genotypes with desirable levels of micronutrients are tested in several locations across target environments to evaluate their genotype-by-environment interaction and stability (Bouis and Saltzman 2017). Many countries in SSA including Uganda, Malawi, Ghana, Mozambique, Kenya, and Ethiopia have recently embarked on biofortification programs. These efforts have contributed to release new orange-, yellow-, and cream-fleshed sweetpotato varieties with emphasis on the former (OFSP), due to its provitamin A content.

To ensure all year-round availability of storage roots of sweetpotato in temperate regions, where production is limited to summer season, roots are cured and stored after harvesting. Careful handling and packing is also necessary to ensure long-term storage. Roots are immediately cured at ~30 °C at relative humidity of 85–90% after harvesting with proper ventilation for 4–7 days. This process facilitates the healing of bruises and wounds that occur during harvesting. Cured roots are then stored at 13–15 °C at 85–95% relative humidity making storage roots available for marketing up to 12 months (Truong et al. 2018).

2.5 The Need for Biotechnology

Biotechnological tools are needed to complement conventional breeding approaches especially for traits whose conventional breeding efforts have not produced the desired results. This assertion is highly manifested in breeding for diseases and pest resistance as well as development of varieties for different end users regarding domestic, commercial, and industrial purposes. Sweetpotato is a polyploid, outcrossing species with high levels of heterozygosity and numerous mating incompatibilities, making its breeding a complex endeavor.

Since the beginning of sweetpotato breeding efforts in 1930, there have been significant improvements in fungal, bacterial, and nematode resistance, β-carotene and anthocyanin levels, yields, storage ability, root size, and starch characteristics. Considerable work continues on insect and virus resistance, processing qualities, and finding regionally adapted cultivars which match consumer preferences (Truong et al. 2018). A significant need still exists for genomic research focused on understanding the inheritance of economically important traits to develop more efficient breeding strategies to facilitate crop improvement. In 2014, the Genomic Tools for Sweetpotato Improvement (GT4SP) project, funded by the Bill & Melinda Gates Foundation (BMGF), addressed this issue and enabled substantial progress in understanding the inheritance of traits of economic importance. These traits include root knot nematode resistance, storage root α- and β-amylases activity, storage root flesh color, and dry matter, starch, and β-carotene contents.

Weevil resistance remains a challenge, and it is still unclear if conventional breeding can result in 100% weevil-resistant varieties, despite many years of work to date focused on breeding for resistance to *Cylas* and *Euscepes* weevils. Several approaches have been suggested to achieve virus and weevil resistance including transgenic approaches, because of the limited progress observed with conventional breeding (Grüneberg et al. 2015). Breeding for weevil resistance requires a

better understanding of the insect's biology and its interaction with plants and environments.

Breeding for desired quality attributes in sweetpotato is challenging because traits of economic importance are often correlated, either positively or negatively. Identifying haplotypes which control traits of economic importance will help in facilitating selection both at the phenotypic and molecular levels. Mapping complex traits is by far the most expensive but also an important approach to identifying functional variants. Some of the quantitative trait loci (QTL) discovered for traits of economic importance were hypothesized to be associated with important candidate genes in sweetpotato which could be targeted in improving cell wall structure, texture, and flavor aside nutritional quality attributes. These could be achieved through the use of biotechnological approaches.

3 Genetic Resources and Sources of Health-Related Genes

3.1 Origin of Sweetpotato and Its Available Germplasms

Ipomoea batatas, the scientific name for the cultivated sweetpotatoes, is a member of the *Ipomoea* genus and Convolvulaceae family. There are more than 600 species described as belonging to the *Ipomoea* genus. Due to the large size of this genus, sweetpotato and its close relatives were gathered in 14 taxa series Batatas. *Ipomoea batatas* is a self-incompatible allogamous species, and its genome is classified as a hexaploid ($2n = 6x = 90$), although earlier studies have revealed the presence of tetraploids ($2n = 4x = 60$) (Austin 1988). Thus, sweetpotato is a polyploid species, and its genetic nature makes genetics and segregation ratios rather complicated.

Since 1985, the International Potato Center (CIP) has conducted breeding studies and developed new sweetpotato varieties. A cooperation between CIP and the International Board for Plant Genetic Resources (IBPGR) was initially established through the collection of extensive amounts of germplasm, focusing on cultivars of sweetpotato and wild *Ipomoea* species throughout Latin America, the region with the most diversity. The collections from the Asian Vegetable Research and Development Center (AVRDC) and International Institute of Tropical Agriculture (IITA) would have been centralized at CIP, which currently houses over 3520 accessions in total. The number of genetic resource collections around the world has expanded, particularly at the national levels, even though a current compilation of germplasm collections is not yet accessible.

The Plant Genetic Resources Conservation Unit (PGRCU), belonging to the United States Department of Agriculture – Agricultural Research Service (USDA-ARS), located in Griffin, Georgia, is responsible for maintaining the US sweetpotato germplasm collection. This genebank contains sweetpotato clonal propagules that are kept as in vitro cultures in addition to maintaining a broad collection of *Ipomoea* species. Hexaploid *I. batatas* clones have been collected over a long period of time, frequently in cooperation with numerous national and international programs and

organizations. The genetic bases for continuous research and advancement in sweetpotato breeding have been provided by this collection.

Based on an analysis of morphological characters of sweet potato and wild *Ipomoea* species, the center of origin of *I. batatas* is thought to lie somewhere between the Yucatan Peninsula in Mexico and the mouth of the Orinoco River in Venezuela (Austin 1988). The highest diversity in Mesoamerica recently revealed by molecular markers indicates that Mesoamerica is an important center of diversity and probably a center of origin, due to the wealth of wild relatives of sweet potato (Huang and Sun 2000). The sweetpotato was brought to Western Europe from the Caribbean after the first voyage of Columbus in 1492. In the sixteenth century, Portuguese explorers took the sweet potato to Africa, India, Southeast Asia, and the East Indies, where it was directly transmitted. The plant was carried by Spanish trading galleons from Mexico to the Philippines. Recent studies assessing genetic diversity using AFLP suggested that prehistoric introductions into Oceania may have originated from the Mesoamerican sweet potato through natural dispersal (Rossel et al. 2001).

Multiple origins have been hypothesized from two separate autopolyploidization events from an ancestral farmed sweetpotato that shares *I. trifida*. According to a recent study (Muñoz-Rodríguez et al. 2018), *I. trifida* was the only progenitor of cultivated sweetpotato, being originated from autopolyploid events. Other studies have suggested that the $B_1B_1B_2B_2B_2B_2$ genome of sweetpotato, which was formed by crossing a tetraploid and a diploid followed by whole genome duplication, has an allo-autohexaploid origin (Yang et al. 2017). The same study claimed that the tetraploid progenitor of cultivated sweetpotato was unknown but assumed the diploid progenitor to be *I. trifida*. According to a newly created genetic map, strict allo-autohexaploid nature cannot exist because no preferential pairing has been observed and haplotypes have shown hexasomic inheritance in an F_1 population (Mollinari et al. 2020).

3.2 Gene Pools and Wild Relatives

It is challenging to genetically improve sweetpotato by conventional plant breeding because of its polyploid nature, genetic complexity, high level of variability in terms of flower production, and incompatibility. The creation of more genomic resources would improve the development of efficient and effective marker-assisted breeding techniques and aid efforts to unravel the genetic causes of phenotypic diversity.

There are 12 species in the *Ipomoea* section *Batatas* whose chromosomes numbers are multiples of 15. The only known hexaploid species in the genus is *I. batatas*, which has sizeable storage roots. All other species into this section are diploid or tetraploid (Austin 1988). The diploid *I. trifida* seems to be the closest wild relative of the cultivated sweetpotato, according to several molecular genetic investigations. In relation to the origin of cultivated sweetpotato, two hypotheses have been provided. Between the Yucatan peninsula and the Orinoco basin, according to Austin's reports from 1988, a natural hybridization between *I. trifida* and *I. triloba* could have produced the ancestors of the sweetpotato. Kobayashi (1983), on the other hand,

hypothesized that *I. trifida* constitutes an autopolyploid complex with ploidy levels ranging from diploid to hexaploid and that domesticated sweetpotato was originated from this group.

It has been suggested that the domesticated sweetpotato, *I. batatas* (6×), is a product of interspecific hybridization between *I. trifida* (2×, 4×, 6×), *I. littoralis* (2×), and *I. leucantha* (2×) (Austin 1988). *I. trifida* and *I. tabascana* have been suggested as the species most closely related to the tetraploid crop (4×). The presumed closest related taxa to sweetpotato are placed in the secondary gene pool in accordance with the gene pool concept of Harlan and de Wet (1971), and because of ploidy incompatibility with the cultivated species: wild forms of *I. batatas* (4×), *I. trifida*, *I. littoralis*, and *I. tabascana* (Jarret and Austin 1994). *I. cordatotriloba*, *I. cynanchifolia*, *I. grandifolia*, *I. lacunosa*, *I. leucantha*, *I. ramosissima*, *I. splendorsylvae*, *I. tenuissima*, *I. tiliacea*, and I. *triloba* are among the species categorized as tertiary wild relatives (Jarret and Austin 1994).

Three novel interspecific hybrids were created by crossing *I. batatas* (L.) Lam. with *I. hederacea* Jacq, *I. batatas* (L.) Lam. with *I. muricata* (L.) Jacq, and *I. batatas* (L.) Lam. with *I. lonchophylla* J. M. Black, as an example of an improved hybridization technique. The storage roots of the first two hybrids were normal, a notable improvement over those of the interspecific *Ipomoea* hybrids that are currently in existence. These three novel interspecific hybrids were genetically compared to three hybrids naturally occurring, namely, diploid, tetraploid, and hexaploid species of the *Ipomoea* section Batatas, using amplified fragment length polymorphism (AFLP) markers. These three novel interspecific hybrids can be used as a bridge to transfer foreign genes from wild species to domesticated species, according to cluster analysis of AFLP bands, which revealed that new interspecific hybrids were closely related to domesticated sweetpotato (Cao et al. 2014). However, due to variances in ploidy levels and inability of wild species to develop storage roots, wild material cannot yet be efficiently used. Further research is needed on certain features and heterosis effects that do not exist in farmed sweetpotato to use wild relatives practically.

4 Classical and Molecular Genetics and Breeding

4.1 Genetics of Health-Related Genes and Breeding Objectives

Classical genetics and conventional breeding principles have enabled the study and improvement of sweetpotato traits which would otherwise not have been possible. Conventional breeding of sweetpotato was initiated in the USA in the early 1900s, relying mainly on plant introductions from the tropics and favorable mutations. With the introduction of sweetpotato flowering under long day conditions, a major breakthrough was realized, which led to the transition from mutation-based breeding to the true seed sexual production-based breeding. A procedure for sweetpotato breeding, based on recurrent selection methods for improvement of maize and other crop species considers the use of polycross and recurrent selection as an effective

way to combine favorable alleles from parental genotypes. This procedure was especially important in sweetpotato for long-term improvement of traits with low heritability.

Population improvement and product development have slowly been performed in sweetpotato due to the complex nature of the crop combined with the constantly changing consumption patterns and environmental conditions. Genotypes that thrive under drought and saline conditions, for example, may be good parents. However, due to the low heritability of drought tolerance and lack of efficient selection strategies, production of drought-tolerant sweetpotato cultivars has been difficult. Although conventional breeding has empowered the improvement of qualities and traits in other crops, this has remained difficult in sweetpotato. Some issues involving sweetpotato conventional breeding can be attributed to male sterility, hexaploid nature, and self-incompatibility.

The breeding objectives of most sweetpotato programs include high yield, tolerance to local biotic and abiotic stresses, and nutritional quality, including traits such as starch, dry matter, and β-carotene contents, as well as amylases, iron, and zinc contents. Specific biotic stresses are dependent on the region where the breeding program is located. Abiotic constraints include low levels and unpredictable rainfall patterns, drought, and low soil fertility. In SSA, the sweetpotato virus disease (SPVD), which is a combination of two different viruses, namely, sweetpotato feathery mottle virus (SPFMBV) and sweetpotato chlorotic stunt virus (SPCSV), is the most common viral stress. *Alternaria* blight, which is a fungal infection, and sweetpotato weevils are both common especially in the eastern parts of Africa. These are the core breeding objectives in those regions.

Increasing emphasis is being focused on "hidden hunger," which is a result of micronutrient deficiencies and not necessarily a low intake of calories. Vitamin A and iron deficiencies affect 50% and 60% of children in SSA, respectively (Mwanga et al. 2021). These are the major causes of anemia and micronutrient deficiency globally. Conventional breeding of sweetpotato has been challenging in this sense. Mass selection has been primarily employed to select resistant clones to advance to the next generation. To do this, a large population of sweetpotato genotypes is assembled and screened for selection of parents and clones for the next breeding cycle.

Phenotyping large populations is challenging and hinders quick progress in terms of selection. A very useful tool for the analysis of health-related traits in sweetpotato is near-infrared spectroscopy (NIRS). It is a nondestructive technique that does not require the use of chemical reagents and has the advantage of being able to screen for numerous constituents with a single scan. Wavelength bands between 780 and 2500 nm penetrate the sample, and light is absorbed selectively according to the sample's molecular composition. Other modern high-throughput phenotyping tools can help with increasing heritabilities and providing higher selection accuracies.

4.2 Molecular Genetics and Inheritance Studies

A large potential exists in the application of DNA-based markers for the improvement of crops. In sweetpotato, the technologies are relatively new and have only

recently initiated to have practical applications into its breeding programs. Previously, most of the studies used dominant gel-based markers, which are relatively easy to score, or microsatellite markers, that have low abundance in the genome (Yada et al. 2017) and only capture limited genetic information. Recently, advances in next-generation sequencing methods and bioinformatics resources have enabled a large number of single nucleotide polymorphism (SNP) markers to be developed and utilized in sweetpotato (Shirasawa et al. 2017; Wadl et al. 2018). SNPs are the most abundant DNA polymorphisms in the genome and can therefore be more readily utilized in a cost-effective genotyping platform. Because they are spread throughout the genome, the genetic information that they capture is massive.

Genetic markers linked to QTL that have been effectively utilized for cultivar development include the QTL for several major crops where the favorable allele presented a large and consistent effect easily measured and fixed by standard breeding procedures. This is more difficult in sweetpotato because of ploidy, high heterozygosity, and high degree of deleterious mutations which are difficult to flush out in asexually propagated species. Maintaining heterozygosity is therefore paramount to keep these mutations at bay, the downside of this being that a typical backcrossing breeding scheme to incorporate a single gene is difficult to achieve.

Few molecular approaches have been exploited in breeding for sweetpotato traits. For SPVD, two recessive genes, *spfm1* and *spcsv1*, were identified for sweetpotato SPFMV and SPCSV from a biparental cross between 'Tanzania' and 'Bikilamaliya' (Mwanga et al. 2002). They used AFLP and random amplified polymorphic DNA (RAPD) markers to map these QTL. In the advent of high throughput next-generation genotyping, it is now possible to identify and utilize SNP markers for genetic studies in sweetpotato. One of the major advantages of SNPs in polyploids is the increased informativeness at a given segregating locus. The development of modern genotyping protocols (Davey et al. 2011), like genotyping-by-sequencing or restriction-site-associated DNA (RAD)-seq, and their optimizations for highly heterozygous and hexaploid sweetpotato, together with a complete reference genome of *I. trifida*, have enabled the evaluation of SNP markers throughout the genome of hexaploid sweetpotato.

Using the previously developed full-sib populations derived from the crosses 'New Kawogo' × 'Beauregard' (NKB), 'Beauregard' × 'Tanzania' (BT), and their reciprocals (TB), together with the new genomic tools developed under the GT4SP and SweetGAINS projects, the core resources needed to bring sweetpotato into the genomics era have been obtained. The new breeding tools include the following: (i) two fully sequenced and annotated diploid lines (*I. trifida* and *I. triloba*) used as reference genomes for cultivated sweetpotato; (ii) a sequencing-based genotyping platform for highly heterozygous hexaploid sweetpotato, *GBSpoly*, with supporting bioinformatics tools (Kuster et al. 2021); (iii) three high-density genetic linkage maps of hexaploid sweetpotato – BT, TB, and NKB; (iv) dosage-dependent SNP calling, phasing, and linkage mapping algorithms for autopolyploids; (v) statistical methods for QTL analysis and linkage mapping in polyploids, implemented in both R packages *QTLpoly* (Da Silva Pereira et al. 2020) and *MAPpoly* (Mollinari et al. 2020), respectively; and (vi) a breeding program database, named *Sweetpotato Base* (Morales et al. 2022).

5 Genetic Diversity Analysis

5.1 Phenotype-Based Diversity Analysis

Diversity can be measured in a variety of ways and helps populations adapt to shifting surroundings. Morphological or molecular markers have been utilized in a number of studies to analyze natural, landrace, or breeding populations of sweetpotato. For instance, a study was reported using morphological characteristics and electrophoretic banding patterns of total proteins and esterase to characterize 1939 Peruvian collections out of 5000 sweetpotato accessions available at the CIP (Huamán et al. 1999). Vargas et al. (2018) evaluated 95 accessions and two commercial cultivars morphologically, and distances were estimated by means of multicategory variables, the data being consequently clustered by the Tocher method. The descriptors that contributed more than 60% of diversity involved the following: leaf size, general leaf profile, immature leaf color, petiole pigmentation, predominant branch color, branch secondary color, stem length, cortical thickness, predominant periderm color, and periderm color intensity.

A phenotypic analysis using morphological and chemical assessments was performed based on 16 factors to complete the study. Italian accessions were found to be extremely close to South American germplasm based on the molecular findings; however, they were subclustered into two groupings. Furthermore, based on similar physical traits and molecular fingerprints, the population's genetic structure analysis revealed that one of the two Italian genotype groups may have descended from South American accessions. Overall, the genetic markers and agronomic parameters used in combination were successful in distinguishing or clumping the sweetpotato genotypes in accordance with their geographic origin or phenotypic descriptors. Breeders and farmers might both use this knowledge to find and safeguard commercial types (Palumbo et al. 2019).

5.2 Molecular Marker-Based Assessment Diversity

According to recent studies on diversity assessment using molecular markers, the largest diversity of sweetpotato was identified in Central America, which supported the idea that this region is the principal center of variability and most likely the sweetpotato's origin (Huang and Sun 2000; Zhang et al. 2000). This origin and its relationships with other species have been studied using different molecular techniques, such as restriction fragment length polymorphism (RFLP), RAPD, microsatellites or simple sequence repeat (SSR), AFLP, and other systems including SNP. Below, we will present some studies on genetic diversity in sweetpotato based on a series of such techniques.

5.2.1 Randomly Amplified Polymorphic DNA (RAPD)

Among eight *Ipomoea* species from section Batatas and 26 accessions of sweetpotato from Peru, Philippines, and the USA, 15 RAPD primers were employed

in this study. The South Pacific and Peruvian sweetpotato lines were readily distinguished in the hexaploid *I. batatas* by a total of 56 polymorphic bands. The species that are most closely related to the farmed hexaploid *I. batatas* are *I. tabascana* and *I. trifida*. These results confirm the value of RAPD markers in assessing sweetpotato genetic diversity and identifying taxonomic and evolutionary links in *Ipomoea* (Jarret and Austin 1994).

5.2.2 Inter-Simple Sequence Repeat (ISSR)

Using ISSR markers, Zhang et al. (2014) evaluated 240 sweetpotato accessions from different germplasms in China. The mean genetic similarity coefficient, Nei's gene diversity, and shared allele distance were each 0.73, 0.32, and 0.27, respectively. Based on the results of Neighbor-Joining (NJ) clustering analysis and STRUCTURE, the 240 accessions could be classified into six subgroups and five subpopulations, and clear genetic links between the tested accessions were found. The marker-based NJ clustering and population structure of the examined sweetpotato germplasm revealed no clear assignment pattern corresponding to flesh color or geographical ecotype. Between white and orange-fleshed sweetpotato accessions, AMOVA found a slight but significant difference. Sweetpotato accessions from the Southern summer-autumn sweetpotato region, the Yellow River Basin spring and summer sweetpotato region, and the Yangtze River Basin summer sweetpotato region all showed small but noticeable differences.

5.2.3 Amplified Fragment Length Polymorphism (AFLP)

Using AFLP markers, 69 cultivars of sweetpotato from four different regions of Latin America were fingerprinted from CIP germplasm (Zhang et al. 2000). The diversity distribution was revealed by mean similarity, multidimensional scaling (MDS), and AMOVA. Central America has the largest genetic variety, whereas Peru-Ecuador had the lowest. These findings are in line with the theory that Peru-Ecuador should be regarded as a secondary center of diversity, with Central America serving as the primary center of diversity and most likely the origin of the sweetpotato.

Bruckner (2004) used AFLP markers to assess the genetic diversity of 75 sweetpotato accessions from CIP, Lima, Peru, and PGRCU, Griffin, USA. AMOVA and main coordinate analysis were applied to the data of 183 polymorphic bands in order to draw the conclusion that the collection had many clusters. Using ten different primer combinations, AFLP analysis of 75 sweetpotato accessions produced 202 distinct polymorphic bands, which allowed for the separation of those accessions. The AFLP analysis results showed that the genetic diversity among the germplasm accessions was relatively modest and that there were few genetic distances between regions.

5.2.4 Simple Sequence Repeat (SSR)

In order to assess the genetic diversity among 112 cultivars in Burkina Faso and to create a core collection, 30 morphological characters and 30 SSR markers were used. While 28 SSR markers were helpful in differentiating the 112 accessions and

identifying 5 duplicates, eight morphological characters were able to distinguish the accessions and identify 11 duplicates. When comparing the results of the two methods, it was found that the phenotypic data produced a high diversity coefficient of 0.73 while the SSR markers produced a moderate diversity coefficient of 0.49. The outcomes demonstrate no relationship between the two methods. Utilizing SSR-based data, a core collection was created, and eight discriminative phenotypic characteristics will be utilized to identify cultivars (Koussao et al. 2014).

Using ten SSR primers, the genetic diversity of 40 sweetpotato accessions from ICAR-CTCRI and CIP was examined. For each SSR and accession, the existence of bands was assessed, and the data were then analyzed using principal coordinate analysis. The sweetpotato cultivars were divided into seven primary clusters using the polymorphic SSR loci, which revealed different relationships among them. Jaccard similarity coefficients between 0.5 and 1.0 in cluster analysis indicated considerable genetic diversity. AMOVA was used to analyze the data, which revealed that the majority of the diversity (82.17%) was found among populations and among the *I. batatas* within the populations (17.83%) (Nair et al. 2017).

Rahman et al. (2023) studied the genetic diversity of 20 sweet potato germplasms originating from Bangladesh, CIP, Philippines, Taiwan, and Malaysia with 20 SSR markers for germplasm characterization and utilization. A total of 64 alleles were generated using the 20 primers from the 20 samples, with locus IBS97 having the highest number of alleles (5), whereas locus IbU33 had the fewest alleles (2). The alleles varied in size from 105 (IbU31) to 213 base pairs (IBS34). The Polymorphism Information Content (PIC) values for the loci IbL46 and IBS97 varied from 0.445 to 0.730. IBS97 has the highest number of effective alleles (3.704), compared to an average of 2.520. The average Shannon's diversity index (H) was 1.003, ranging from 0.673 in IbU3 to 1.432 in IBS97. The value of gene flow (Nm) varied between 0.000 and 0.005, with an average of 0.003, whereas genetic differentiation (Fst-values) ranged between 0.901 and 1.000. Total 20 genotypes were classified into two groups in the UPGMA dendrogram, classified as group "A" with 16 genotypes and the remaining classified as group "B" with 4 genotypes. As a result, SSR markers could be successfully evaluated for the genetic relationships among the sweet potato accessions used and generated a high level of polymorphism.

Liu et al. (2023) studied the genetic diversity of 617 sweetpotato accessions from China using 30 SSR primers. Based on the population structure analysis, these accessions were grouped into three groups, Group 1, Group 2, and Group 3, containing 228, 136, and 253 accessions, respectively. The results were obtained via phylogenic analysis and principal coordinate analysis (PCoA). Of the three groups, Group 2 showed the highest level of genetic diversity and were mainly distributed in low-latitude regions. The accessions from South China exhibited the highest level of genetic diversity, which assists the hypothesis that Fujian and Guangdong were the earliest regions where sweetpotato was brought to China. Analysis of molecular variance (AMOVA) showed that there were significant genetic differences between different groups.

5.2.5 Start Codon Targeted (SCoT) Polymorphisms

Using ten SCoT primers, 40 sweetpotato accessions obtained from ICAR-CTCRI and CIP had their SCoT polymorphisms examined. There were created 128 bands in all, 75 of which were polymorphic among the accessions. With a product size range of 200 bp to 2 kb and a polymorphism of 56.5%, the primer SCoT 11 produced the greatest number of bands (25), whereas the primer SCoT 21 produced the fewest bands (7). The genotypes of the sweetpotato lines employing the SCoT marker were grouped into three major clusters as a result of the hierarchical cluster analysis. Nearly all of the ICAR-CTCRI released varieties fell into cluster I, while the CIP high-carotene clones were divided into clusters II and III. These findings suggested that the SCoT marker system is useful for identifying polymorphism and separating variability among sweetpotato lines (Nair et al. 2016).

5.2.6 SNP

Using a genotyping-by-sequencing (GBS) protocol optimized for highly heterozygous and polyploid species, GBSpoly, Wadl et al. (2018) reported the diversity study of 417 USDA sweetpotato accessions coming from eight broad geographical regions (Africa, Australia, Caribbean, Central America, Far East, North America, Pacific Islands, and South America). The accessions were divided into four genetic groups by population structure using Bayesian clustering analysis (STRUCTURE) with 32,784 polymorphic SNPs. Indicating a significant degree of mixed ancestry, the population structure analysis was supported by a neighbor-joining cladogram and principal components analysis based on a pairwise genetic distance matrix of the accessions. Based on the origin of accessions, pairwise Fst values between major geographic areas ranged from 0.017 (Far East – Pacific Islands) to 0.110 (Australia – South America).

Feng et al. (2018) identified the diversity and relationship of different polyploid types in sweet potato and *I. trifida* using Restriction-site Associated DNA sequencing (RAD-seq). A total of 38,605 RAD tags containing 832,204 SNPs were identified. These tags were annotated with five public databases, and approximately 11,519 tags were mapped to functional genes in different pathways. Based on the SNP genotype, the results of phylogenetic analysis confirmed that cultivated sweet potato is closely related to wild species, *I. trifida*.

Using 662 parents from the International Potato Center (CIP) global breeding program covering Peru, Uganda, Mozambique, and Ghana, a low-density, very informative set of SNP markers has been developed to be used for routine quality assurance and control (QA/QC) (Gemenet et al. 2020b). Segregation of the selected 30 SNPs was evaluated in a recombinant breeding population using 282 offspring of the selected parents. Genotypes were reproduced in vitro, in a screen house, and in the field, and it was confirmed that the selected SNP set identified relatively similar similarities as a dense SNP set of 10,159 markers. The study also allowed us to assess the global breeding population structure of CIP and the impact of some of the most devastating stressors, such as sweet potato virus disease, on the management of genetic variation. These results provide insight into the future implementation of genomic selection in sweet potato.

5.3 Relationship with Geographical Distribution

To represent both domesticated gene pools, *I. batatas* (329 landraces) was sampled across its distributional range from Mexico to Peru by nuclear and chloroplast microsatellite markers. Both markers maintained the presence of two geographically restricted genepools, corresponding to accessions from the Peru–Ecuador region of South America (Southern genepool) and accessions from the Caribbean and Central America region (Northern genepool). According to the tripartite hypothesis of sweet potato introduction into Oceania (Oceania corresponds here to the entire insular region between Asia and the Americas, including Polynesia, Micronesia, and Melanesia), the three lines introduced different genepools into different areas. The distributional ranges of *I. trifida* (40 accessions) and *I. triloba* (15 accessions) were sampled. It has been hypothesized that one accession of *I. tabascana* is a cross between *I. batatas* and *I. trifida* (McDonald and Austin 1990).

Further research strongly suggested that *I. batatas* may comprise actual wild populations with lower ploidy levels in addition to the hexaploid cultigen, from which cultivated forms would have been domesticated. In addition, recent research suggests that the domestication of the sweetpotato may have had two separate origins. Cultivated landraces have two distinct geographically clustered chloroplast lineages, with the other two landraces being found in northwestern South America. These differentiation patterns were supported by nuclear microsatellite markers (Roullier et al. 2011).

Six geographical regions and a total of 23 countries that produce sweetpotato were used to analyze the genetic links among 74 sweetpotato accessions using a total of 71 polymorphic RAPD molecular markers. The genotypes from Central or the Caribbean and South America formed two distinct clusters, according to MDS approach. East African variations stood out from other traditional kinds from South America and Oceania due to their distinctive traits from other traditional varieties. These findings lend support to the proposed theory on the genesis and spread of the sweetpotato and showed that the main source of diversity likely contains two different gene pools. It is suggested that, rather than variants from South America, the spread of the sweetpotato from its origin may have mostly involved types from Central America or the Caribbean. There is evidence that new gene pools may be developing in Africa and Asia as a result of hybridization and environmental adaption (Gichuki et al. 2003).

According to Khoury et al. (2015), the richest areas for sweetpotato's closest wild relatives are Mesoamerica and the extreme Southeastern United States, which stretch from the Central United States to Argentina. The adaptation of designated species to temperature, precipitation, and edaphic features varies among them and in comparison to the crop, and the majority of species have also exhibited significant intraspecific variation. Because 79% of the species were deemed high priority for crop genetic resources and in ex situ conservation systems, breeders and researchers' access to these species was insufficient. Additional taxonomic study, characterization, and evaluation of germplasm, as well as developing methods to get over

obstacles to introgression with wild species, are required in order to mobilize the genetic resources for improved conservation of sweetpotato's wild relatives.

6 Molecular Mapping of Health-Related Genes and QTLs

6.1 Molecular Markers: Types, Evolution of Molecular Markers

Individuals within a species will present variations at the DNA level among themselves. These genetic variations are the source for the development of molecular markers, which provide genetic information at specific loci for individuals within a population. As such, molecular markers are employed in plant breeding for germplasm characterization, gene introgression tracking, genetic variability assessment, genetic and QTL mapping, and cultivar protection. In diploids, molecular markers application is straightforward, but for polyploids there are quite a few hurdles for their usage. The major hurdle is the large number of genotypic classes. Thus, most markers can only be scored as dominant (single-dose markers), and allele dosage or multiallelic information is lost. Although several advances have been attained in the polyploid statistics field, molecular and statistical methods still have a gap among them to efficiently estimate genotypic classes. As a hexaploid species, the application of molecular markers in sweetpotato breeding is still hindered by such difficulties.

The introduction of molecular markers in sweetpotato genetic studies started in the early 1990s. Jarret et al. (1992) used RFLP markers to study the phylogenetic relationships among 13 *Ipomoea* species with different ploidy levels. The marker analysis showed that the diploid *I. trifida* and wild tetraploids are phylogenetic close to *I. batatas*. RAPD markers were the first ones used to study genetic linkage in sweetpotato (Thompson et al. 1997). Gichuki et al. (2003) fingerprinted 74 sweetpotato varieties with RAPD markers to infer the relationship among genotypes from different geographic sources. South American and Central American varieties formed two separate clusters, probably indicating two centers of diversity. Additionally, the results suggested that new gene pools might have been evolving in Africa and Asia due to hybridization and local environments adaptation.

The first sweetpotato genetic map with considerable coverage was constructed using AFLP markers (Kriegner et al. 2003). Rossel et al. (2001) employed AFLP markers to investigate the dispersal of sweetpotato to Oceania by comparing the relationships among samples from Peru, Ecuador, Mexico, Philippines, and eight Oceania countries. The results showed the occurrence of gene flow between Mexico and Oceania, and little relationship with Peru-Ecuador genotypes. Sweetpotato genetic diversity was also greater in Oceania, and the authors hypothesized a nonhuman dispersal from Mesoamerica to Oceania. AFLP markers were also used in disease resistance studies.

SSRs, also called microsatellites, are highly polymorphic and are scattered throughout the genome. Only in diploids, though, they can be analyzed as codominant markers. The first SSRs for sweetpotato were developed in 1994 (Jarret and Bowen 1994), and since then the development and usage of these markers in

sweetpotato only increased (Zhang et al. 2001; Roullier et al. 2011; Roullier et al. 2013; Yada et al. 2017). Zhang et al. (2001), using SSR markers, observed a distinct geographical pattern within sweetpotato Latin America gene pool. The authors proposed that Mesoamerica is the primary center of diversity and probable center of origin, while Peru-Ecuador region would be the secondary center of diversity. SSR markers also indicated the existence of two domesticated sweetpotato gene pools, including the Northern and Southern genotypes, which suggests two domestication events (Roullier et al. 2011). In addition, these markers also showed that sweetpotato most probably has an autopolyploid origin, sharing an ancestor with *I. trifida*. Considering the high genetic diversity present in sweetpotato, the species seem to not have suffered a severe bottleneck during domestication (Roullier et al. 2013). Associations of SSR loci with desirable sweetpotato agronomic characteristics were also identified (Yada et al. 2017).

In polyploids, SNPs can have multiple doses of a nucleotide allele. Being a hexaploid, sweetpotato can have up to seven different genotypic classes: six copies of the reference nucleotide (*AAAAAA*), one to five copies of the alternative nucleotide (*AAAAAa*, *AAAAaa*, *AAAaaa*, *AAaaaa*, and *Aaaaaa*), and six copies of the alternative nucleotide (*aaaaaa*). Inferring allele dosage from next-generation sequencing data is now possible in sweetpotato (Mollinari et al. 2020) by the usage of software such as SuperMASSA (Serang et al. 2012).

6.2 Genetic Linkage Maps

In general, the construction of genetic linkage maps is performed to generate a representation of the genome and to provide an appropriate structure for QTL mapping. This construction is dependent on the genetic structure and size of the population and the genome complexity. In sweetpotato, genetic linkage maps are constructed using F_1 segregating populations derived from biparental crosses between highly heterozygous individuals. It occurs because this crop has self-incompatibility and does not generate inbred or homozygous lines, which are explored in many of the experimental crosses used for mapping in plants.

In this sense, several F_1-based genetic linkage maps have been published for sweetpotato to date. The first one was proposed by Ukoskit and Thompson (1997), who used 196 RAPD markers for constructing two low-density genetic linkage maps based on 76 F_1 progenies. Two maps were obtained separately for each parent ('Vardaman' and 'Regal') using 25 and 20 single-dose (simplex) markers segregating in a 1:1 ratio, following the two-way pseudo-testcross approach. The final maps were obtained incorporating the double-dose (duplex) and triple-dose (triplex) markers, increasing the total number of polymorphisms to 48 ('Vardaman') and 46 ('Regal'). With the nonsimplex markers, authors identified homologous groups in both maps, improving the informativeness.

More recently, Shirasawa et al. (2017) developed a high-density genetic linkage map based on 142 progenies derived from self-pollination of the 'Xushu 18' cultivar. A total of 94,361 SNP markers were identified after filtering for coverage depth

(≥10) and missing data (<0.25), of which 36,590 (38,8%) were double-simplex (expected segregation of 1:2:1). However, because the number of reads was insufficient to distinguish two genotypic classes (*AAAAAa* and *AAAAaa*), the authors coded the data as dominant loci, with an expected segregation of 3:1. After filtering to Mendelian segregation, the genetic linkage map was constructed using 28,087 (29,8%) SNPs, which generated 96 LGs (90 + six small extra groups) and covered a total distance of 33,020.4 centiMorgan (cM). On average, each LG presented a length of 344 cM and 293 SNPs.

Although all referenced genetic maps have contributed to represent the sweetpotato genome and confirmed its hexasomic segregation, they were limited on its characterization. Most of them used the two-way pseudotestcross approach to construct one map for each parent, which was proposed for diploid species. Moreover, this strategy does not allow the construction of an integrated map containing multiple markers with different segregation types, which should improve the genetic characterization. In addition, most of maps were built using a two-point approach, which does not consider multiple markers simultaneously to the map construction.

In this sense, Mollinari et al. (2020) proposed the first integrated multilocus genetic map to simultaneously estimate recombination fractions and linkage phases for a complex polyploid. The method was applied for a sweetpotato F_1 progeny consisting of 315 individuals, which was originated from a cross between the 'Beauregard' and 'Tanzania' cultivars. An ultradense genetic map consisting of 30,684 SNP markers (~2046 SNPs per LG) was built, covering 2708.3 cM (180.6 cM per LG; 11.3 SNPs/cM) and containing the complete hexaploid haplotypes for the mapping population. This comprehensive map showed that most of the meiotic configurations (73.3%) were bivalents and clearly evidenced a hexasomic-bivalent inheritance in sweetpotato, which should promote stability to the allelic transmission along the generations.

6.3 QTL Mapping

Traits selected by breeding programs are usually quantitative and strongly influenced by environment. Generally, breeders need to evaluate a lot of candidate genotypes in multiple locations and years to select the best ones, which takes a long time and lots of resources. One way to accelerate this process is to use DNA-based markers (i.e., without environmental influence) genetically associated with quantitative traits to perform selection, based on the marker-assisted selection (MAS). Different approaches can be used to identify markers associated with traits. Here, we will present the QTL mapping, focusing on the studies developed for sweetpotato.

QTL mapping, i.e., association between the genotype (DNA-based markers) and the phenotype (trait expression), can be performed using different statistical methods such as single marker analysis (SMA) and interval mapping (IM) based models (Lander and Botstein 1989). Except for SMA, all these methods use a genetic linkage map to search for QTL, considering that mapping population is diploid and originated from inbred parents. Although sweetpotato is an autohexaploid and

outcrossing species, these methods were applied in previous studies of the crop. Currently, QTL mapping can be specifically performed for a complex autopolyploid and outcrossing species (Da Silva Pereira et al. 2020), and we will discuss the application of this approach in sweetpotato. First, we will briefly describe the previous studies.

For the first mapping QTL in sweetpotato, SMA detected nine markers/QTL associated with root-knot nematodes (RKN) resistance, being seven QTL for one parent ('Tanzania') and two QTL for the other parent ('Beauregard'). For 'Tanzania', four QTL showed positive additive effects on RKN resistance and explained approximately 21% of the phenotypic variation. For 'Beauregard', the two QTL had positive additive effects on RKN resistance and explained 6% of the phenotypic variation (Cervantes-Flores et al. 2008b). Authors have also used IM and composite interval mapping (CIM) on previous genetic maps (Cervantes-Flores et al. 2008a) and confirmed the same results of SMA.

QTL mapping has also been done for quality traits in this crop. Cervantes-Flores et al. (2011), who were first to map QTL in this context, used CIM in two previous genetic linkage maps (Cervantes-Flores et al. 2008a) to identify 13 (15–24%), 12 (17–30%), and 8 (17–35%) QTL for dry matter, starch, and β-carotene, respectively. The same traits were investigated by Yada et al. (2017), who used another mapping population (Yada et al. 2015) and SMA to detect four, six, and eight SSR markers associated with dry matter, starch, and β-carotene. These markers, which could directly be associated with traits or in linkage disequilibrium with QTL, explained overall 15.8, 32.3, and 37.8% of the phenotypic variation, respectively.

Although all these studies were important and made progress, they were limited to map QTL in sweetpotato, due to its genetic complexity. Recently, Da Silva Pereira et al. (2020) proposed a comprehensive QTL mapping for sweetpotato using an extension of MIM for a complex autopolyploid. Considering a previous integrated linkage map (Mollinari et al. 2020), the authors detected a total of 13 QTL for eight yield-related traits, with the number of QTL per trait ranging from one to four. These QTL explained up to 55% of the variation, being that both parents ('Beauregard' and 'Tanzania') contributed with alleles to increasing the trait means in the mapping population.

Using the same mapping population of Mollinari et al. (2020) and Da Silva Pereira et al. (2020), Gemenet et al. (2020a) identified two major QTL related to starch, β-carotene, and their respective correlated traits, dry matter and flesh color. The authors used the extension of MIM proposed by Da Silva Pereira et al. (2020), which enabled to evidence that both QTL presented a pleiotropic effect to reducing starch and increasing β-carotene in genotypes carrying a haplotype from the 'Beauregard' parent.

Mwanga et al. (2021) and Oloka et al. (2021) also used the extension of MIM for autopolyploids (Da Silva Pereira et al. 2020) to detect QTL in sweetpotato. The former mapped a total of six QTL (two per trait) for β-carotene and other important traits such as iron (Fe) and zinc (Zn) contents, and the QTL explained 65.8%, 51.0%, and 23.5% of the phenotypic variation, respectively. The latter identified one major QTL (58.3% of the phenotypic variation) for RKN resistance, which would be a putative candidate to perform MAS for this important disease.

In summary, QTL mapping has been used in different populations (backgrounds) and for many traits in sweetpotato to date. Great progress has been achieved, particularly with the recent studies which used molecular markers and statistical methods specifically developed for an autopolyploid and outcrossing species. Next steps in QTL mapping should remain in the extension of the methods to account for multiple traits or environments simultaneously, enabling the investigation of pleiotropic effects and linkage as well as the interaction between QTL and environments. Certainly, the future results of these approaches will be helpful to improve our understanding regarding genetic architecture of the traits, providing valuable information for MAS into sweetpotato breeding.

6.4 Association Mapping

Association mapping is a powerful approach to identifying QTL that requires a large number of polymorphisms and overcomes the need for biparental populations. Diversity panels are used instead, allowing for the exploitation of historical recombinant events, relying on linkage disequilibrium (i.e., nonrandom associations at different loci) to find associations between markers and traits. It identifies alleles controlling complex traits and presents higher resolution than linkage analysis, but its effectiveness depends on marker density, sample size, population structure, and relatedness.

There are two main approaches of association mapping: genome-wide association studies (GWAS) and candidate gene analysis. The selection of approach depends on the availability of markers. While candidate gene association uses markers previously chosen based on their locations and on previous QTL studies, GWAS test associations across the entire genome, using individuals that are densely genotyped. Association mapping has been successfully used in major diploid crop species, where the availability of genomic tools and resources is large. However, in autopolyploid species, such as sweetpotato, genomic analysis has been challenging for a long time, mainly due to the complex genomes. Autopolyploid species do not resemble the meiotic chromosome behavior of diploids (Yamamoto et al. 2020), so allele dosage information is required to effectively map complex traits using all the allelic information of the genotypes and population.

The first attempts in polyploids were based on AFLP or SSR markers, which are used as dominant markers, and scored according to the presence or absence of bands. Advances in DNA sequencing via next-generation sequencing technology have enabled the usage of SNP, the most abundant markers in the genome. In autopolyploids, allele dose estimation for array- and mass spectrometry-based SNP data dates back to 2012 (Serang et al. 2012).

In autopolyploids, GWAS has evolved from using diploid models, where marker data are "diploidized," to models accounting for allele dosage. GWASpoly (Rosyara et al. 2016) is the first R package designed for GWAS with SNP in autopolyploids using the classic Q + K method, which takes into account both population structure (Q) and cryptic relatedness (K). GWASpoly can conduct single marker association

tests using different models of gene action. The modified Q + K model was evaluated and validated in GWASpoly using simulated data and a tetraploid potato diversity panel with nulliplex (*aaaa*), simplex (*Aaaa*), duplex (*AAaa*), triplex (*AAAa*), and quadriplex (*AAAA*) states.

In sweetpotato, a study aimed to determine whether the genetic mapping of polyploids could be possible when there is no accurate allele dose information. The applicability of the low-coverage NGS was confirmed using real data from autohexaploid sweetpotato. The authors performed an association study using an analogous approach to the "general" option presented in GWASpoly, but not including the kinship matrix. Three peaks on homologous group 6 were detected, with simplex dominant phenotypic inheritance mode, where all heterozygotes were equivalent to one of the homozygotes.

An association study using 46,788 SNPs was performed to map resistance to southern root-knot nematode (SRKN; *Meloidogyne incognita*) race SP2 in hexaploid sweetpotato using multiple-dose markers. One peak was found on chromosome 7 and another on chromosome 3, both loci presenting a simplex state. DNA markers were developed to screen for nematode resistance using the SNPs present on the peaks. The accuracy of selecting SRKN-SP2-resistant plants using the two markers was approximately 70% (Obata et al. 2022).

7 Marker-Assisted Breeding for Health-Related Traits

7.1 Germplasm Characterization

Germplasm characterization of varieties and breeding populations has been essential for crop improvement, and diversity has long been seen as vital for rational management and use of crops. A crucial step in breeding for improved nutrient content is the characterization of quality steps. Genetic diversity analysis is key for deciphering the extent of genetic variation available between and within germplasm collections.

Molecular markers are used in complementarity with phenotypic markers when trait expression is environmentally unstable. The most predominant marker system used in the past few decades were the SSR markers or microsatellites. These markers are highly polymorphic, and codominant, and can easily be detected on high-resolution gels. Marker systems keep on evolving, and those based on genomic SNPs have gained center stage in germplasm characterization, due to their amenability for high-throughput sequencing and detection.

7.2 Limitations and Prospects of Marker-Assisted Breeding

MAS is an approach that identifies and selects genotypes based on markers rather than phenotypes per se. MAS can improve the selection efficiency since it (i) can be performed early on seedlings, (ii) is not affected by environmental conditions, (iii) can identify recessive alleles controlling the trait of interest, (iv) can perform gene

pyramiding, and (v) allows selecting for traits with low heritability. However, the use of MAS in routine breeding programs presents some limitations, including (i) the cost: MAS is more expensive than conventional breeding. Even though markers are cheap, the initial cost can be high; (ii) the degree of linkage between the marker and the target gene: recombination between the marker and the target gene may occur as far as the marker is from the gene; (iii) markers are, sometimes, population-specific; and (iv) QTLs have distinct effect in different genetic backgrounds: QTL effects are not always consistent across environments, and the magnitude of their effects depends upon QTL by environment interaction.

Marker-assisted backcrossing breeding (MABCB) is considered an applicable form of MAS. It is an effective approach to introgress major genes from a donor parent into the elite and improved variety, while recovering the recurrent parent genome through multiple backcross generations, minimizing linkage drag. The MABCB involves the usage of markers to select for target loci, minimizing the segment containing the loci of interest and accelerating recovery of the recurrent parent during backcrossings by using markers. Selection strategies for MABCB for a gene in diploids find three generations of backcrossings to be sufficient to recover a high proportion of the recurrent parent, saving two to four backcross generations in the transfer of a single locus. Such an approach, however, is hardly applicable to sweetpotato breeding due to its strong inbreeding depression and self-incompatibility, as well as the polyploid and complex nature of its genome.

7.3 GWAS and Genomic Selection (GS)

GWAS and genomic selection (GS) are effective approaches to investigating marker-trait associations. Both approaches are based on genome-wide markers. GWAS use genotypic and phenotypic information from diversity panels to estimate marker effects across the genome. The analysis is based on linkage disequilibrium and can identify novel QTL and genes for the trait of interest.

Instead of identifying specific associations, GS uses all genome-wide markers simultaneously to predict the performance of individual candidates for selection. The idea is that QTLs will be in linkage disequilibrium with some markers because the marker information is dense enough to do so. The selection is based on genomic estimated breeding values (GEBVs) obtained from a predictive model that integrates genotypic and phenotypic data of a training population. The selection of superior individuals based on GEBVs leads to a shorter breeding cycle since it is not necessary to wait for late progeny generations to phenotype the traits of interest. It promotes acceleration of process and reduction of costs, enabling to effectively deliver increased genetic gain per unit time and cost.

Breeders of polyploids and clonally propagated species such as sweetpotato have additional questions regarding the usage of GS models in their breeding programs. The first one is related to the inclusion or not of nonadditive (dominance and epistasis) genetic effects in the models. Modeling nonadditive genetic effects in sweetpotato, whether diploidized or with dosage information, it was found that the

models without nonadditive effects were comparatively better in prediction accuracy than the models including nonadditive effects (Gemenet et al. 2020c). A second question was whether the dosage information significantly improves accuracy. The authors stated the relative importance of considering dosage is dependent on trait architecture. The general rule is that the more complex the trait, the more important dosage information is.

The GS strategy using low-depth sequencing data applied by the blueberry breeding program at the University of Florida may be helpful for a range of other polyploid species. Their approach simplified the allele dosage information in auto-polyploids and resulted in similar prediction accuracies compared with more refined scenarios (Ferrão et al. 2021). The authors stated the inclusion of allele dose information on GS models might improve the accuracy of the GEBVs by considering the additive effect of dose and the potential dominance effects.

A combined GS plus de novo GWAS model was proposed to combine the results of GWAS using the GS training population in rice (Spindel et al. 2016). Briefly, the model uses the results of the GWAS as part of the GS model, with no need for extra datasets. The authors stated the significant SNPs identified from the GWAS would be relevant to the population under selection and that the method is more suitable for traits with one or more medium-large QTL segregating in the population. Since statistical models have evolved for autopolyploids, GS plus de novo GWAS could also be applied to sweetpotato breeding programs.

8 Genomics-Aided Breeding for Health-Related Traits

8.1 Health-Related Functional Genomics Resources

Access to large-scale gene expression profiling datasets from a wide range of species has revealed genes controlling certain health traits and those in response to abiotic and biotic stresses. For sweetpotato, a reference genome sequence from its wild *I. trifida* has recently been made available (Wu et al. 2018), enabling more genomic resources deployment to address sweetpotato genome complexity (Wadl et al. 2018; Wu et al. 2018; Da Silva Pereira et al. 2020). This has aided the understanding of genetic architecture and control of β-carotene and starch in sweetpotato as well as unraveling genes in sweetpotato carotenogenesis (Gemenet et al. 2020a) and those responding to abiotic stress such as drought (Lau et al. 2018). So far, a few studies have reported genes responding to health-related (HR) traits such as carotenoids and starch (Wu et al. 2018; Gemenet et al. 2020a; Mwanga et al. 2021).

Insights into genes involved in regulation of the pathways for carotenoid and starch/carbohydrate metabolic pathways are well characterized in many crops, but still generally lacking in sweetpotato. Besides, until recently (Gemenet et al. 2020a), the molecular mechanisms driving the association between starch and carotenoids metabolic pathways in sweetpotato were studied independent of each other.

Genome redundancy exhibited by *Brassica* species provides a model to study the evolutionary fate of multicopy genes and the effects of polyploidy in economically

important crops. Significant work has been done unraveling the mode of action of phytoene synthase (PSY), as the catalyst of the first committed reaction of the carotenoid biosynthetic pathway and the main bottleneck in the carotenoid pathway (Cazzonelli and Pogson 2010). In *Arabidopsis thaliana*, a single *PSY* gene (*AtPSY*) is responsible for regulating phytoene synthesis in all tissues. However, in most plant species, the *PSY* gene family seems to be composed of two or three homologous genes. For example, carrots (Just et al. 2007) have two PSY encoding genes while in other crops, such as maize and tomato (Fraser et al. 1999), three *PSY* homologs have been reported. The three *PSY* copies in grasses have subfunctionalized and provided a fine tine control of carotenogenesis in response to various developmental and external cues (Li et al. 2009).

While some isoforms participate in the biosynthesis of carotenoids in chloroplast-containing photosynthetic tissues, others participate in the production of carotenoids in nonphotosynthetic tissues of the fruit (tomato *PSY1*), seed endosperm (maize *PSY1*), or the root (tomato, maize, and rice *PSY3*). The *Arabidopsis PSY* gene is expressed in virtually all tissues, including both photosynthetic and non-photosynthetic (Rodríguez-Villalón et al. 2009), and shows high rates of coexpression with the rest of the genes involved in the carotenoid pathway and in the supply of their isoprenoid precursors (Meier et al. 2011). The Orange protein regulates carotenoid accumulation by posttranscriptionally regulating PSY, promoting the formation of carotenoid-sequestering structures, and also preventing carotenoid degradation. The Orange protein contributes to homeostasis regulation, improving plant tolerance to abiotic stress (Park et al. 2016).

Orange flesh color in sweetpotato has been correlated with high β-carotene content. Logical control points/loci for the accumulation of β-carotene in OFSP were identified by Wu et al. (2018), providing targets for MAS of OFSP varieties. Three SNPs in gene homologs encode key carotenoid biosynthetic enzymes, including PSY (*itf03g05110*), phytoene desaturase (PDS; *itf11g08190*), and ζ-carotene isomerase (Z-ISO; *itf04g12320*), in a set of sweetpotato parents composed of orange- and white-fleshed accessions. Upregulation of the *PSY* SNPs in orange-fleshed 'Beauregard' storage roots compared to other toot types suggested their involvement in conferring the orange flesh in 'Beauregard' storage roots. Furthermore, gene expression profiling of root development in 'Beauregard' showed biased expression of the "orange" alleles of the SNPs in *PSY* at late (at 40 and/or 50 days after transplanting) development stage of the storage roots, proposing expression-based increment of carotenoid accumulation. However, one SNP associated with orange flesh in *PDS* showed a weak but significant biased expression of the "orange" allele, at an early stage of storage root development (20 DAT), and a not significant in later stage of storage roots development (50 DAT). This showed the importance of the expression of specific PDS alleles before storage root initiation in the development of orange flesh.

QTL and differential gene expression analyses of β-carotene and starch loci, in a biparental population, may perhaps be one of the most comprehensive studies where the two traits have been studied together unraveling the genetic basis for the negative association between β-carotene and starch (Gemenet et al. 2020a). The study

contained 315 F_1 genotypes generated from an OFSP cultivar, 'Beauregard', and a white fleshed cultivar, 'Tanzania'. Enabled by the availability of a sweetpotato reference genome (Wu et al. 2018) and new linkage and QTL mapping methods for polyploids (Mollinari et al. 2020; Da Silva Pereira et al. 2020), four correlated traits, namely, dry matter, starch, β-carotene, and flesh color, were placed on QTLs based on an integrated genetic map of the 15 LGs of sweetpotato (Gemenet et al. 2020a). Variation for the traits in the mapping population was explained by a major QTL colocalized on LG3 and LG12. At LG12 QTL, both parents contributed with major alleles presenting similar allelic effects on the traits; however, the OFSP parent, 'Beauregard', contributed the major allelic effect to traits at the LG3 QTL.

Dry matter, starch, β-carotene, and flesh color are traits affected by the same haplotypes which act in completely opposite directions, i.e., same haplotypes participating in decreasing the means for dry matter and starch were also responsible for increasing the means for β-carotene and flesh color. This explains the observed negative association between starch and β-carotene in sweetpotato, a phenomenon also seen in cassava (Rabbi et al. 2017). Given that the two parents are contrasting for these traits, it appears that interaction of alleles in QTL on LG12 with those QTL on LG3 determines the accumulation or lack of accumulation of β-carotene in the storage roots of sweetpotato. Furthermore, the rate-limiting gene in the carotenoid biosynthesis, *PSY* gene (*itf03g05110*), was reported to sit in between the two colocalized QTL peaks at LG3, one at 2,994,719 bp (for β-carotene and flesh color) and the other at 3,185,578 bp (for dry matter and starch). Moreover, gene expression profiling datasets from developing roots showed carotenoid accumulation in both storage and fibrous roots in 'Beauregard' but not 'Tanzania'.

Similar expression levels of the starch gene *sucrose synthase* were found (*SuSY*; *itf03g05100*) during the early days of root development (10–20 days) for both parents but persist only in 'Tanzania' after 50 DAT. 'Beauregard' has a maturity period of 60 DAT while 'Tanzania' matures in 150 days. Moreover, *PSY* gene and *SuSY* genes are located within a reach of 12.2 kb with no intervening genes (Gemenet et al. 2020a). *Orange* (*Or*) gene homolog (*itf12g24270*) was found near (5.7 kb) the QTL peak associated with β-carotene on LG12. The *Or* gene has been reported to act on and regulate *PSY* allowing modification of amyloplasts into chromoplasts in several species including OFSP cultivars (Park et al. 2015). Together, this analysis identified logical control points for the accumulation of β-carotene in sweetpotato and loci that provide targets for MAS.

Starch is the most important carbohydrate in storage root in sweetpotato accounting for about 75% of human caloric intake. The composition, size, and shape of starch granules contributes to sweetpotato eating quality (Kitahara et al. 2017). Starch is primarily synthesized and accumulated as starch grains (SGs) in amyloplasts, which are also modified to store carotenoids in starchy organs such as seeds of wheat, rice, barley, and maize, as well as potato tubers, sweetpotato, and cassava roots. A mutually exclusive relationship exists between carotenoid accumulation and starch granule development in tobacco floral nectaries and carrot roots, suggesting that enhanced carotenogenesis serves as a developmental signal that directs the transition from amyloplasts to chromoplasts (Kim et al. 2010). However, storage of carotenoids in modified amyloplasts results in carbon competition for both starch

biosynthesis and carotenoid biosynthesis, which leads to the negative association reported in several crops including sweetpotato (Yada et al. 2017; Gemenet et al. 2020a).

8.2 Other Functional Genomics Resources

Expression profiling has become a valuable tool to investigate how an organism responds to environmental changes and can define both tolerant and sensitive responses. Profiles of plant response to environmental extremes are expected to lead to useful regulators in biotechnological approaches to improve stress tolerance and new tools for studying regulatory genetic circuitry. Alterations in the plant transcriptome may lead to further insights into effect of abiotic stress to plant physiology. De novo transcriptome studies have been carried out for hexaploid sweetpotato under abiotic stresses such as drought and salt (Lau et al. 2018; Arisha et al. 2020b; Zhou et al. 2022), and heat (Arisha et al. 2020a) in different *Ipomoea* species.

Even though sweetpotato tolerates abiotic stress, both drought salinity and heat stress still affect its productivity. Transcriptome studies have led to the discovery of key genes, gene networks, and transcription factors involved in single or regulatory roles in multiple abiotic stress responses in sweetpotato. While some genes respond to multiples stressors, others are stress specific. For instance, aquaporins (Arisha et al. 2020a) and Late embryogenesis proteins (LEAs), including late embryogenesis abundant protein and dehydrin, which are considered as a part of stress-protecting protein families in sweetpotato, can facilitate water uptake especially during drought. LEAs were highly upregulated in the water stress experiments (Lau et al. 2018). These proteins play a protective role under osmotic stress conditions, and homeostasis of proteins and nucleic acids.

Several stress-inducible genes have been identified in sweetpotato through transcriptomics, but their function within the molecular mechanisms for stress response and tolerance still needs to be deciphered. For example, the production of the phytohormone abscisic acid (ABA) causes stomatal closure, induces the expression of stress-responsive genes (Lau et al. 2018; Arisha et al. 2020b; Zhou et al. 2022), and plays a key responsive element binding factors (ABFs)-regulatory role in multiple abiotic stresses responses (Wang et al. 2019). Expression of *IbABF4* was induced by ABA and several environmental stresses including drought, salt, and heat shock (Wang et al. 2019). However, how they work is unknown. Further research trials tailoring cultivars for abiotic tolerance and enhanced yield by overexpression of the genes in sweetpotato should be increased all over the globe with the premier goal of more crops.

8.3 Gene Editing and Transgenics

The need to edit genomes keeps on evolving in many crops with sweetpotato not being an exception. Genome sequences have been developed for several crops, and

their availability and advances in genome editing approaches have made way to breed for almost any given desirable trait (Jaganathan et al. 2018). Examples of genome editing technologies include zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced palindromic repeats (CRISPR/Cas9). CRISPR/Cas9-mediated genome editing is a powerful technology that facilitates the genetic modification of a number of species. In plant breeding of crops with valuable traits such as antioxidant-rich purple tomatoes, omega-3 fatty acid-enriched oil crop, starch altered potato, and high yielding rice, CRISPR/Cas9 system has been successfully used (Jaganathan et al. 2018). Previous studies have demonstrated that CRISPR/Cas9 technology is an effective tool for the improvement of starch qualities in sweetpotato and breeding polyploid root crops (Wang et al. 2019).

Some of the opportunities for applying nanotechnology in sweetpotato improvement lie in the areas of genetic improvement (Mukhopadhyay 2014) such as delivery of genes and drug molecules to specific sites at the cellular level in plants and animals. In addition, nanoarray-based technologies for gene expression in plants, to overcome stress and development of sensors and protocols for its application in precision farming, are prospective tools that could be considered for the improvement of sweetpotato crop. Furthermore, management of natural resources and early detection of pathogens and contaminants in food products also present good opportunities for improvement (Mukhopadhyay 2014).

Efforts were made to change the two starch components amylose and amylopectin. Amylose-free genotypes were generated using a transgenic approach (Kimura et al. 2001). Through gene stacking means, transgenic sweetpotato plants that produce both carotenoid and anthocyanin in the same storage root were developed by overexpressing the *IbOr* gene, involved in accumulation of carotenoids, or the *IbMYBI* gene, a transcription factor involved in the biosynthesis of anthocyanin (Kang et al. 2017). Using the CRISPR/Cas9 technology, altering the amylose/amylopectin ratio through selective knockout of *IbGBSSI* or *IbSBEII* was demonstrated (Wang et al. 2019).

9 Role of Bioinformatics as a Tool

9.1 Gene and Genome Databases

In the past decades, completion of the *Arabidopsis* and rice genome sequences and the continuing sequencing of other economically important nonmodel crop plants have been carried out. These efforts have provided an unprecedented amount of genome sequence data for large-scale genomics studies and analyses. For example, discovery and identification of novel genes, comparative genomics, and functional genomics can now be performed.

The ease of access and organization of the data primarily determines the efficient utilization of these large data sets in databases. Several plant databases have been set up to maintain various data types including genomic sequence, annotation, analyses

expressed transcript assemblies and analyses, and gene expression profiles from microarray studies. Several plant genomic databases have been developed. However, while some are inclusive, others are relational species-specific databases containing substantial amounts of data from one species such as tje *Arabidopsis* Initiative Resource (TAIR: http://www.arabidopsis.org/) database for the model plant *Arabidopsis thaliana*.

Increased genome projects have enabled the creation of databases with various related species. One example is the Solanaceae Genomics Network database (SGN; http://solgenomics.net/), comprising the reference genomes of tomato, potato, pepper, and *Solanum pennelli*; the draft genomes of *Solanum pimpinellifolium*, *Nicotiana benthamiana*, and *N. tabaccum*; and the genomes of two inbred lines of *Solanum lycopersicum*. The Sweetpotato Genomics Resource (SGR; http://sweetpotato.uga.edu) was recently developed to host the genome sequences and RNA-seq gene expression data of *I. trifida* and *I. triloba,* wild ancestors of the cultivated hexaploid sweetpotato. The Sweetpotato Genomics Resource serves as a reference for genetic improvement of the cultivated hexaploid sweetpotato (*I. batatas*) (Wu et al. 2018). Moreover, the *Ipomoea* Genome Hub is an integrated web-based database for *Ipomoea* species and hosts the haplotype-resolved genome sequence of hexaploid *I. batatas* (cv. Taizhong6) (Yang et al. 2017). Despite the efforts, the genome assembly is still incomplete and contains a considerable amount of redundancy and misassemblies, including erroneous assembly of haplotypes and therefore not of much use as a reference for genome-enabled breeding.

9.2 Comparative Genome Databases

The genome consists of a unique gene list controlling the specific phenotype and its interaction with the environment; however, this changes during evolution. New genes can arise when evolution tinkers with old ones but can also be caused by the natural selection. Variations, caused by natural selection, are encoded in an organism's genome, and provide clues to their genetic divergence from a common ancestor. The inference of variations between species based on gene inventories compositions opens the door to the rich branch of comparative genomics. The current innovations and speedy progress in sequencing technologies have substantially improved whole genome data providing scientists with a large dataset for orthologous-sequence comparisons.

Several databases have been developed and enabled comparative analysis of various sequenced plant genomes. The National Center for Biotechnology Information (NCBI) genome database provides a large suite of online resources for biological information and data, including the genebank nucleic acid sequence database and the PubMed database of citations and abstracts published in life science journals, giving access to genomic information for science advancement (Sayers et al. 2010). NCBI uses an integrated database retrieval system website termed Entrez, reaching a diverse set of databases that together contain billions of records. Entrez acts as the analysis and retrieval resources for the data in GenBank and other biological data

fields (Sayers et al. 2010). The NCBI Assembly database (www.ncbi.nlm.nih.gov/assembly/) gives stable access and data tracking for genome assembly data and may accommodate a range of assembly structures, including sets of an unordered contig or scaffold sequences depending on the database model. NCBI serves as a valuable genomics resource for model crops; however, it is not the case for sweetpotato (Table 2). The limited availability of sweetpotato genomic resources is further demonstrated by its absence in other integrative databases like the Phytozome (https://phytozome-next.jgi.doe.gov/), a comparative hub for plant genome and gene family data and analysis, and Ensembl Plants (http://plants.ensembl.org/).

The new generation of comparative genomics offers a powerful aid to identify biologically active regions of a genome based on the observation that sequences that hold key functions are frequently conserved between cross-species sequences, and thus can be distinguished from nonfunctional surrounding sequences. Therefore, comparative genomics is seen in this sense, as equivalent to evolutionary genomics. Orthology gene search using comparative genomics data reveals gene expansions and contractions that a species has undergone during evolution and can advance our understanding of the process of adaptation. A good illustration is found in the lineage-specific expansion of genes encoding sporamin proteins in both *I. trifida* and *I. triloba*. Sporamin (Kunitz-type trypsin inhibitor – KTI) is analogous to patatin in potato and a key storage protein in sweetpotato storage roots; the latter plays an important role in storage, defense, and development (The Potato Genome Sequencing Consortium 2011).

Comparative analysis offers an unprecedented understanding of gene duplications, new gene functions, and the impact of gene duplication on genome evolution. Through comparative analyses, whole genome triplicate events (WGT) were found to have

Table 2 Comparisons of genomic resources available in the NCBI database for sweetpotato compared to other common crops

Species	Assembly	Genome	Genes	Nucleotides	SRA	BioSample	Bioproject
Ipomoea batatas	3	1	14,097	464,681	5418	4761	156
Triticum aestivum	48	1	183,127	2,973,037	115,887	58,710	1424
Zea mays	85	16	74,430	5,320,607	130,769	104,306	5302
Oryza sativa	105	1	57,069	2,933,812	118,091	120,019	7054
Sorghum bicolor	15	1	46,948	1,365,256	31,267	26,293	5785
Solanum lycopersicum	17	1	57,119	938,027	26,706	25,248	1228
Solanum tuberosum	28	1	73,099	538,023	15,802	15,222	455
Saccharum officinarum	12	9	29,637	534,355	1509	1243	162
Manihot esculenta	9	1	34,354	358,460	8756	26,632	130

occurred in an ancient ancestor of the *Ipomoea* lineage specifically after the divergence with Solanaceae. Enrichment of WGT genes ontologies were suggestive of regulatory functions. For instance, two genes associated with storage root development, sweetpotato knotted-like homeobox genes, *IBKN2* (orthologous to *itf01g32840*) and *IBKN3* (*itf01g32840*), were duplicated in the ancestral *Ipomoea* WGT event (Wu et al. 2018). These genes also had higher expression level in 'Beauregard' storage roots compared with fibrous roots and were upregulated during the development of storage roots suggesting that the WGT contributed to additional gene copies that function in storage root development (Wu et al. 2018). Moreover, the strategy for comparative genomics between plants and distantly related relatives has been used to identify functional genes among different plants species and also helps researchers study genes annotation in newly sequenced plant species.

Access to the data is provided through a genome browser incorporating many specialist interfaces for different data types, and through a variety of additional methods for programmatic access and data mining. Bioinformatics tools used to compare the genome sequences, and the RNAs, proteins, and gene annotations from different organisms, are constantly evolving to deal with the exponential production of sequenced genomes driven by advances in sequencing technology, and therefore becoming more comprehensive and user-friendly.

9.3 Gene Expression Databases

Gene expression data from transcriptome experiments (microarray and RNA-seq datasets) have been archived in public repositories such as NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) and the EBI ArrayExpress (AE; https://www.ebi.ac.uk/arrayexpress/). Furthermore, the DNA DataBank of Japan (DDBJ) recently started a repository called the Genomic Expression Archive (GEA: https://www.ddbj.nig.ac.jp/gea/). These databases are useful resources for the functional interpretation of genes but have been separately maintained and may lack RNA-seq data, while the original sequence data are available in the Sequence Read Archive (SRA). Millions of RNA sequence sequences have been generated by sweetpotato experiments involving different abiotic and biotic conditions (Lau et al. 2018; Wu et al. 2018; Gemenet et al. 2020a; Arisha et al. 2020b), developmental stages (Gemenet et al. 2020a), and health traits such as β-carotenoid (Wu et al. 2018; Gemenet et al. 2020a), anthocyanin (Xiong et al. 2022), and starch (Qin et al. 2021) which are all available at the NCBI SRA. Several known stress-responsive transcription factors (TFs) in *Arabidopsis* and rice have been used to correctly predict stress-responsive TFs in other plant species, including sweetpotato (Zheng et al. 2016).

9.4 Other "Omics" Databases

Plant proteome research is an exciting and essential frontier of integrative biology. Proteome analysis in different plant systems has been carried out using various tissue

samples such as leaves, root, flower, and fruits, among others. Analysis of the proteome component of an organism offers structural, functional, and network scale insights; however, most existing databases focus on DNA/RNA data or specific gene families, with less emphasis on protein structure, function, and variability. Furthermore, high-throughput mass spectrometric techniques augment understanding of proteins and their expression patterns under various conditions. The Protein DataBase, PPDB, is a whole Plant Proteome DataBase dedicated for *Arabidopsis thaliana* and maize (*Z. mays*). Supplemented with plastid and mitochondrial genomes, protein-encoding gene models in *Arabidopsis*, maize, and rice, as assembled, respectively, in http://www.arabidopsis.org/, http://www.maizesequence.org/index.html, and http://rice.plantbiology.msu.edu/, are uploaded in PPDB and are linked to each other via a BLAST alignment (Sun et al. 2009). Thus, every predicted protein in both species can be searched for experimental and other information (even if not experimentally identified). The PPDB stores experimental data from in-house proteome and mass spectrometry analysis, curated information about protein function, protein properties, and subcellular localization. Importantly, proteins are particularly curated for possible (intra) plastid location and their plastid function. Protein accessions identified in published *Arabidopsis* (and other Brassicacea) proteomics papers are cross-referenced to rapidly determine previous experimental identification by mass spectrometry.

Protein-protein interactions (PPIs) play key roles in diverse biological processes, and therefore it is necessary to understand the interaction details of plant PPIs. Experimentally validated or predicted PPI data have thus become increasingly available in diverse plant species. Protein interaction details can be annotated at various levels by integrating bioinformatics algorithms, and then compiled into user-friendly databases. PlaPPISite (http://zzdlab.com/plappisite/index.php) is a comprehensive, high-coverage, and interaction details-oriented database for 13 plant interactomes (Yang et al. 2020). The source species of interolog templates, GO annotations, subcellular localizations, and gene expression similarities are provided to facilitate the reliability assessment of predicted PPIs.

The Plant Metabolic Network (PMN) constitutes the PlantCyc, a database containing biochemical pathway and their catalytic enzymes, compounds, and genes from various plant species (Hawkins et al. 2021). A greater number of pathway diagrams in PlantCyc were manually extracted from the plant literature but are either supported by experimental evidence or are based on expert hypotheses. Enzymes and genes are experimentally verified and assigned to pathways with evidence codes to provide information about data quality.

The stage for genomics, transcriptomics, and proteomics, coupled with bioinformatics and biostatistics, are gaining momentum. However, they are still, for the most part, assessed individually with distinct approaches generating monothematic rather than integrated knowledge. In addition, despite a fast-growing number of available plant genomes, available computational resources are poorly integrated and provide only limited access to the underlying data. The PlantsDB gives a good example of a platform that is inclusive of integrative and comparative plant genome research (Spannagl et al. 2007). PlantsDB is constituted from genome databases for

Arabidopsis, Medicago, lotus, rice, maize, and tomato, and it might be leveraged to help sweetpotato research in its data integration process.

10 Final Considerations and Prospects

Sweetpotato cultivation and consumption can distinctively be partitioned based on flesh color. In SSA, selective breeding has been conducted over the years for high dry matter (>25%) in varieties that have white, cream, and yellow flesh colors with higher starch content and have mealy firm texture after cooking. The OFSP varieties with high β-carotene level, but low dry matter content (18–25%), are commercially popular in the USA. They are sweet and have a moist texture after cooking. Overall, in comparison with other tuber crops, tropical sweetpotato varieties comprise higher contents of carbohydrates, various minerals, and more protein estimates than other vegetables. The protein content of sweetpotato is low (~2%), but still higher than cassava and plantain, for example.

Efforts to biofortify sweetpotato have focused on increasing β-carotene content and improving organoleptic qualities of commonly consumed varieties. Continuous efforts have been made to improve the nutritional value of landrace sweetpotato varieties through biofortification, leading to improved carotenoid content. These improvements have been successful through the adoption of advanced breeding techniques, which involve the screening of large number of genotypes for nutritional quality, agronomic traits, yield traits etc., in order to select progenies with the best traits for further breeding (Yada et al. 2017; Gemenet et al. 2020a). Replacement of white-fleshed sweetpotato with OFSP varieties has benefited ~50 million children below 6 years of age at risk of vitamin A deficiency (Low et al. 2017). Furthermore, OFSP clones are being selected for other health traits such as Fe and Zn. Several studies have reported a common set of genetic loci controlling β-carotene, Fe, and Zn contents and highly correlated levels providing a basis to improve more than one trait concurrently (Mwanga et al. 2021).

Purple sweetpotato have attractive purple-red color, high anthocyanin content, high total phenol content, and high antioxidant activity (Steed and Truong 2008). The anthocyanin content in purple sweetpotato is similar to those of anthocyanin temperate fruits, such as blueberries, blackberries, cranberries, and grapes. Moreover, purple sweetpotato is a low-cost source of natural anthocyanin pigments. The PFSP anthocyanins contain various chemical structures mainly cyanidins and peonidins in the form of monoacylation and deacetylation. These acylated forms give advantages to heat and light stability as well as antioxidant activity (Mu et al. 2021), which is an advantage as natural pigments in food additives. Besides, the PFSP varieties have high dry matter content even up to 38.96% compared to the OFSP and a negative correlation between anthocyanin level and water content (Steed and Truong 2008).

Marker-assisted breeding allows assessment of young plants at the seedling stage for multiple traits of interest, greatly reducing costs associated with growing the plants to maturity. This approach is especially valuable for sweetpotato, where the expense of

long-term field evaluation is a major limiting factor in breeding efforts. It is not feasible to conduct backcrossing breeding to introgress simple or oligogenic traits. Furthermore, genomic data provide a foundation to elucidate genetic relationships among parental lines and potentially identify new sources of genetic variation associated with environmental tolerance, pest and disease resistance, and other high-value traits. GS may also facilitate the assessment of hardiness, resistance to emerging diseases and insect pests, and changing consumer preferences. Less expensive genome sequencing, innovative methods for the construction of nucleic acid libraries, improved mapping methodologies, and advanced computational approaches make genomics-based breeding an attractive and powerful option for the improvement of sweetpotato.

Although sweetpotato is a globally important crop, investment in genomic resources to unravel its full genetic potential has been inadequate. In real-world plant breeding, it is a major challenge to pyramid polygenic traits and integrate them with the quest of obtaining high yield and end-user preference simultaneously. In order to dissect the mechanism of inheritance, phenotypic assessment needs to be correlated with a large number of molecular markers spread across genomes using linkage maps or high-density SNP markers across the genome. Transgenic and genome editing also show potential in tackling such HR traits (Okwuonu et al. 2022). Changes to the conventional breeding schemes, such as accelerated breeding schemes (Andrade et al. 2017) and participatory breeding (Gibson et al. 2011) can help with the quick turnaround time for both release and adoption of new healthier sweetpotato varieties.

References

Amankwaah VA (2019) Phenotyping and genetic studies of storage root chemistry traits in sweetpotato. Ph.D. Dissertation, North Carolina State University

Anbuselvi S, Muthumani S (2014) Phytochemical and antinutritional constituents of sweet potato. J Chem Pharm Res 6(2):380–383

Andrade MI, Ricardo J, Naico A, Alvaro A, Makunde GS, Low J, Ortiz R, Grüneberg WJ (2017) Release of orange-fleshed sweetpotato (*Ipomoea batatas* [L.] Lam.) cultivars in Mozambique through an accelerated breeding scheme. J Agric Sci 155(6):919–929. https://doi.org/10.1017/S002185961600099X

Arisha MH, Aboelnasr H, Ahmad MQ, Liu Y, Tang W, Gao R, Yan H, Kou M, Wang X, Zhang Y, Li Q (2020a) Transcriptome sequencing and whole genome expression profiling of hexaploid sweetpotato under salt stress. BMC Genomics 21(1):197. https://doi.org/10.1186/s12864-020-6524-1

Arisha MH, Ahmad MQ, Tang W, Liu Y, Yan H, Kou M, Wang X, Zhang Y, Li Q (2020b) RNA-sequencing analysis revealed genes associated drought stress responses of different durations in hexaploid sweet potato. Sci Rep 10(1):12573. https://doi.org/10.1038/s41598-020-69232-3

Austin DF (1988) The taxonomy, evolution and genetic diversity of sweetpotato and related wild species. In: Report of the 1st sweet potato planning conference. International Potato Center, Lima, pp 27–59

Bouis HE, Saltzman A (2017) Improving nutrition through biofortification: a review of evidence from HarvestPlus, 2003 through 2016. Glob Food Secur 12:49–58. https://doi.org/10.1016/j.gfs.2017.01.009

Bruckner AW (2004) AFLP-based genetic diversity assessment of global sweetpotato (*Ipomoea batatas* (L.) Lam.) germplasm resources: Progress toward the development of a sweetpotato core collection. MS Thesis, North Carolina State University

Cao Q-H, Tang J, Li A, Gruneberg W, Huamani K, Ma D (2014) Ploidy level and molecular phylogenic relationship among novel *Ipomoea* interspecific hybrids. Czech J Genet Plant Breed 50(1):32–38. https://doi.org/10.17221/172/2013-CJGPB

Cazzonelli CI, Pogson BJ (2010) Source to sink: regulation of carotenoid biosynthesis in plants. Trends Plant Sci 15(5):266–274. https://doi.org/10.1016/j.tplants.2010.02.003

Cervantes-Flores JC, Yencho GC, Kriegner A, Pecota KV, Faulk MA, Mwanga ROM, Sosinski BR (2008a) Development of a genetic linkage map and identification of homologous linkage groups in sweetpotato using multiple-dose AFLP markers. Mol Breed 21(4):511–532. https://doi.org/10.1007/s11032-007-9150-6

Cervantes-Flores JC, Yencho GC, Pecota KV, Sosinski B, Mwanga ROM (2008b) Detection of quantitative trait loci and inheritance of root-knot nematode resistance in sweetpotato. J Am Soc Hortic Sci 133(6):844–851. https://doi.org/10.21273/JASHS.133.6.844

Cervantes-Flores JC, Sosinski B, Pecota KV, Mwanga ROM, Catignani GL, Truong VD, Watkins RH, Ulmer MR, Yencho GC (2011) Identification of quantitative trait loci for dry-matter, starch, and β-carotene content in sweetpotato. Mol Breed 28:201–216. https://doi.org/10.1007/s11032-010-9474-5

Da Silva PG, Gemenet DC, Mollinari M, Olukolu BA, Wood JC, Diaz F, Mosquera V, Gruneberg WJ, Khan A, Buell CR, Yencho GC, Zeng Z-B (2020) Multiple QTL mapping in autopolyploids: a random-effect model approach with application in a hexaploid sweetpotato full-sib population. Genetics 215(3):579–595. https://doi.org/10.1534/genetics.120.303080

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12 (7):499–510. https://doi.org/10.1038/nrg3012

Feng JY, Li M, Zhao S, Zhang C, Yang ST, Qiao S, Tan WF, Qu HJ, Wang DY, Pu ZG (2018) Analysis of evolution and genetic diversity of sweetpotato and its related different polyploidy wild species I. trifida using RAD-seq. BMC Plant Biol 18(1):181. https://doi.org/10.1186/s12870-018-1399-x

Ferrão LFV, Amadeu RR, Benevenuto J, de Bem OI, Munoz PR (2021) Genomic selection in an outcrossing autotetraploid fruit crop: lessons from blueberry breeding. Front Plant Sci 12. https://doi.org/10.3389/fpls.2021.676326

Fraser PD, Kiano JW, Truesdale MR, Schuch W, Bramley PM (1999) Phytoene synthase-2 enzyme activity in tomato does not contribute to carotenoid synthesis in ripening fruit. Plant Mol Biol 40 (4):687–698. https://doi.org/10.1023/a:1006256302570

Gemenet DC, da Silva PG, De Boeck B, Wood JC, Mollinari M, Olukolu BA, Diaz F, Mosquera V, Ssali RT, David M, Kitavi MN, Burgos G, Felde TZ, Ghislain M, Carey E, Swanckaert J, Coin LJM, Fei Z, Hamilton JP, Yada B, Yencho GC, Zeng Z-B, Mwanga ROM, Khan A, Gruneberg WJ, Buell CR (2020a) Quantitative trait loci and differential gene expression analyses reveal the genetic basis for negatively associated β-carotene and starch content in hexaploid sweetpotato [*Ipomoea batatas* (L.) Lam.]. Theor Appl Genet 133(1):23–36. https://doi.org/10.1007/s00122-019-03437-7

Gemenet DC, Kitavi MN, David M, Ndege D, Ssali RT, Swanckaert J, Makunde G, Yencho GC, Gruneberg W, Carey E, Mwanga RO, Andrade MI, Heck S, Campos H (2020b) Development of diagnostic SNP markers for quality assurance and control in sweetpotato [*Ipomoea batatas* (L.) Lam.] breeding programs. PLoS One 15(4):e0232173. https://doi.org/10.1371/journal.pone.0232173

Gemenet DC, Lindqvist-Kreuze H, De Boeck B, da Silva PG, Mollinari M, Zeng Z-B, Craig Yencho G, Campos H (2020c) Sequencing depth and genotype quality: accuracy and breeding operation considerations for genomic selection applications in autopolyploid crops. Theor Appl Genet 133(12):3345–3363. https://doi.org/10.1007/s00122-020-03673-2

Gibson RW, Mpembe I, Mwanga ROM (2011) Benefits of participatory plant breeding (PPB) as exemplified by the first-ever officially released PPB-bred sweet potato cultivar. J Agric Sci 149 (5):625–632. https://doi.org/10.1017/S0021859611000190

Gichuki ST, Berenyi M, Zhang D, Hermann M, Schmidt J, Glössl J, Burg K (2003) Genetic diversity in sweetpotato [*Ipomoea batatas* (L.) Lam.] in relationship to geographic sources as assessed with RAPD markers. Genet Resour Crop Evol 50(4):429–437. https://doi.org/10.1023/A:1023998522845

Grüneberg WJ, Ma D, Mwanga ROM, Carey EE, Huamani K, Diaz F, Eyzaguirre R, Guaf E, Jusuf M, Karuniawan A, Tjintokohadi K, Song YS, Anil SR, Hossain M, Rahaman E, Attaluri SI, Somé K, Afuape SO, Adofo K, Lukonge E, Karanja L, Ndirigwe J, Ssemakula G, Agili S, Randrianaivoarivony JM, Chiona M et al (2015) Advances in sweetpotato breeding from 1992 to 2012. In: Low J, Nyongesa M, Quinn S, Parker M (eds) Potato and sweetpotato in Africa: transforming the value chains for food and nutrition security, 1st edn. Oxfordshire, CABI International, pp 3–68

Harlan JR, de Wet JMJ (1971) Toward a rational classification of cultivated plants. Taxon 20(4):509–517. https://doi.org/10.2307/1218252

Hawkins C, Ginzburg D, Zhao K, Dwyer W, Xue B, Xu A, Rice S, Cole B, Paley S, Karp P, Rhee SY (2021) Plant metabolic network 15: a resource of genome-wide metabolism databases for 126 plants and algae. J Integr Plant Biol 63(11):1888–1905. https://doi.org/10.1111/jipb.13163

Huamán Z, Aguilar C, Ortiz R (1999) Selecting a Peruvian sweetpotato core collection on the basis of morphological, eco-geographical, and disease and pest reaction data. Theor Appl Genet 98(5):840–844. https://doi.org/10.1007/s001220051142

Huang JC, Sun M (2000) Genetic diversity and relationships of sweetpotato and its wild relatives in *Ipomoea* series *Batatas* (Convolvulaceae) as revealed by inter-simple sequence repeat (ISSR) and restriction analysis of chloroplast DNA. Theor Appl Genet 100(7):1050–1060. https://doi.org/10.1007/s001220051386

Jaganathan D, Ramasamy K, Sellamuthu G, Jayabalan S, Venkataraman G (2018) CRISPR for crop improvement: an update review. Front Plant Sci 9. https://doi.org/10.3389/fpls.2018.00985

Jarret RL, Austin DF (1994) Genetic diversity and systematic relationships in sweetpotato (*Ipomoea batatas* (L.) Lam.) and related species as revealed by RAPD analysis. Genet Resour Crop Evol 41(3):165–173. https://doi.org/10.1007/BF00051633

Jarret RL, Bowen N (1994) Simple Sequence Repeats (SSRs) for sweetpotato germplasm characterization. Plant Genet Resour Newsl 100:9–11

Jarret RL, Gawel N, Whittemore A (1992) Phylogenetic relationships of the sweetpotato [*Ipomoea batatas* (L.) Lam.]. J Am Soc Hortic Sci 117(4):633–637. https://doi.org/10.21273/JASHS.117.4.633

Just BJ, Santos CAF, Fonseca MEN, Boiteux LS, Oloizia BB, Simon PW (2007) Carotenoid biosynthesis structural genes in carrot (*Daucus carota*): isolation, sequence-characterization, single nucleotide polymorphism (SNP) markers and genome mapping. Theor Appl Genet 114(4):693–704. https://doi.org/10.1007/s00122-006-0469-x

Kang L, Park S-C, Ji CY, Kim HS, Lee H-S, Kwak S-S (2017) Metabolic engineering of carotenoids in transgenic sweetpotato. Breed Sci 67(1):27–34. https://doi.org/10.1270/jsbbs.16118

Khoury CK, Heider B, Castañeda-Álvarez NP, Achicanoy HA, Sosa CC, Miller RE, Scotland RW, Wood JRI, Rossel G, Eserman LA, Jarret RL, Yencho GC, Bernau V, Juarez H, Sotelo S, de Haan S, Struik PC (2015) Distributions, *ex situ* conservation priorities, and genetic resource potential of crop wild relatives of sweetpotato [*Ipomoea batatas* (L.) Lam., *I.* series *Batatas*]. Front Plant Sci 6:251. https://doi.org/10.3389/fpls.2015.00251

Kim JE, Rensing KH, Douglas CJ, Cheng KM (2010) Chromoplasts ultrastructure and estimated carotene content in root secondary phloem of different carrot varieties. Planta 231(3):549–558. https://doi.org/10.1007/s00425-009-1071-7

Kimura T, Otani M, Noda T, Ideta O, Shimada T, Saito A (2001) Absence of amylose in sweet potato [*Ipomoea batatas* (L.) Lam.] following the introduction of granule-bound starch synthase I cDNA. Plant Cell Rep 20(7):663–666. https://doi.org/10.1007/s002990100376

Kitahara K, Nakamura Y, Otani M, Hamada T, Nakayachi O, Takahata Y (2017) Carbohydrate components in sweetpotato storage roots: their diversities and genetic improvement. Breed Sci 67(1):62–72. https://doi.org/10.1270/jsbbs.16135

Kobayashi M (1983) The *Ipomoea trifida* complex closely related to sweet potato. In: Proceedings of the sixth symposium of the International Society for Tropical Root Crops. International Potato Center, Lima, pp 561–568

Koussao S, Gracen V, Asante I, Danquah EY, Ouedraogo JT, Baptiste TJ, Jerome B, Vianney TM (2014) Diversity analysis of sweet potato (*Ipomoea batatas* [L.] lam) germplasm from Burkina Faso using morphological and simple sequence repeats markers. Afr J Biotechnol 13(6):729–742. https://doi.org/10.5897/AJB2013.13234

Kriegner A, Cervantes JC, Burg K, Mwanga ROM, Zhang D (2003) A genetic linkage map of sweetpotato [*Ipomoea batatas* (L.) lam.] based on AFLP markers. Mol Breed 11(3):169–185. https://doi.org/10.1023/A:1022870917230

Kuster RD, Yencho GC, Olukolu BA (2021) ngsComposer: an automated pipeline for empirically based NGS data quality filtering. Brief Bioinform 22(5):bbab092. https://doi.org/10.1093/bib/bbab092

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121(1):185–199. https://doi.org/10.1093/genetics/121.1.185

Lau KH, del Rosario HM, Crisovan E, Wu S, Fei Z, Khan MA, Buell CR, Gemenet DC (2018) Transcriptomic analysis of sweet potato under dehydration stress identifies candidate genes for drought tolerance. Plant Direct 2(10):e00092. https://doi.org/10.1002/pld3.92

Li F, Tsfadia O, Wurtzel ET (2009) The phytoene synthase gene family in the gGrasses. Plant Signal Behav 4(3):208–211

Liu C, Zhao N, Jiang Z, Zhang H, Zhai H, He S, Gao S, Liu Q (2023) Analysis of genetic diversity and population structure in sweetpotato using SSR markers. J Integr Agric. https://doi.org/10.1016/j.jia.2023.02.004

Low JW, Mwanga ROM, Andrade M, Carey E, Ball A-M (2017) Tackling vitamin a deficiency with biofortified sweetpotato in sub-Saharan Africa. Glob Food Secur 14:23–30. https://doi.org/10.1016/j.gfs.2017.01.004

McDonald JA, Austin DF (1990) Changes and additions in *Ipomoea* section *Batatas* (Convolvulaceae). Brittonia 42(2):116. https://doi.org/10.2307/2807625

Meier S, Tzfadia O, Vallabhaneni R, Gehring C, Wurtzel ET (2011) A transcriptional analysis of carotenoid, chlorophyll and plastidial isoprenoid biosynthesis genes during development and osmotic stress responses in *Arabidopsis thaliana*. BMC Syst Biol 5(1):77. https://doi.org/10.1186/1752-0509-5-77

Mollinari M, Olukolu BA, da Pereira GS, Khan A, Gemenet D, Yencho GC, Zeng Z-B (2020) Unraveling the hexaploid sweetpotato inheritance using ultra-dense multilocus mapping. G3 Genes Genomes Genet 10(1):281–292. https://doi.org/10.1534/g3.119.400620

Morales N, Ogbonna AC, Ellerbrock BJ, Bauchet GJ, Tantikanjana T, Tecle IY, Powell AF, Lyon D, Menda N, Simoes CC, Saha S, Hosmani P, Flores M, Panitz N, Preble RS, Agbona A, Rabbi I, Kulakow P, Peteti P, Kawuki R, Esuma W, Kanaabi M, Chelangat DM, Uba E, Olojede A, Onyeka J, Shah T, Karanja M, Egesi C, Tufan H, Paterne A, Asfaw A, Jannink J-L, Wolfe M, Birkett CL, Waring DJ, Hershberger JM, Gore MA, Robbins KR, Rife T, Courtney C, Poland J, Arnaud E, Laporte M-A, Kulembeka H, Salum K, Mrema E, Brown A, Bayo S, Uwimana B, Akech V, Yencho C, de Boeck B, Campos H, Swennen R, Edwards JD, Mueller LA (2022) Breedbase: a digital ecosystem for modern plant breeding. G3 Genes Genomes Genet 12(7):jkac078. https://doi.org/10.1093/g3journal/jkac078

Mu J, Xu J, Wang L, Chen C, Chen P (2021) Anti-inflammatory effects of purple sweet potato anthocyanin extract in DSS-induced colitis: modulation of commensal bacteria and attenuated bacterial intestinal infection. Food Funct 12(22):11503–11514. https://doi.org/10.1039/D1FO02454J

Mukhopadhyay SS (2014) Nanotechnology in agriculture: prospects and constraints. Nanotechnol Sci Appl 7:63–71. https://doi.org/10.2147/NSA.S39409

Muñoz-Rodríguez P, Carruthers T, Wood JRI, Williams BRM, Weitemier K, Kronmiller B, Ellis D, Anglin NL, Longway L, Harris SA, Rausher MD, Kelly S, Liston A, Scotland RW (2018)

Reconciling conflicting phylogenies in the origin of sweet potato and dispersal to Polynesia. Curr Biol 28(8):1246–1256.e12. https://doi.org/10.1016/J.CUB.2018.03.020

Mwanga ROM, Kriegner A, Cervantes-Flores JC, Zhang DP, Moyer JW, Yencho GC (2002) Resistance to sweetpotato chlorotic stunt virus and sweetpotato feathery mottle virus is mediated by two separate recessive genes in sweetpotato. J Am Soc Hortic Sci 127(5):798–806. https://doi.org/10.21273/JASHS.127.5.798

Mwanga ROM, Swanckaert J, da Silva PG, Andrade MI, Makunde G, Grüneberg WJ, Kreuze J, David M, De Boeck B, Carey E, Ssali RT, Utoblo O, Gemenet D, Anyanga MO, Yada B, Chelangat DM, Oloka B, Mtunda K, Chiona M, Koussao S, Laurie S, Campos H, Yencho GC, Low JW (2021) Breeding progress for vitamin a, iron and zinc biofortification, drought tolerance, and sweetpotato virus disease resistance in sweetpotato. Front Sustain Food Syst 5: 616674. https://doi.org/10.3389/fsufs.2021.616674

Nair AGH, Vidya P, Mohan C (2016) Analysis of genetic variability in sweet potato accessions using Start Codon Targeted (SCoT) polymorphism. Int J Biotechnol Biochem 12(2):111–121

Nair AGH, Vidya P, Ambu V, Sreekumar J, Mohan C (2017) Genetic diversity studies in cultivated sweet potato (*Ipomoea batatas* (L.) Lam) revealed by simple sequence repeat markers. Int J Adv Biotechnol Res 7(1):33–48

Neela S, Fanta SW (2019) Review on nutritional composition of orange-fleshed sweet potato and its role in management of vitamin A deficiency. Food Sci Nutr 7(6):1920–1945. https://doi.org/10.1002/fsn3.1063

Nguyen HC, Chen C-C, Lin K-H, Chao P-Y, Lin H-H, Huang M-Y (2021) Bioactive compounds, antioxidants, and health benefits of sweet potato leaves. Molecules 26(7):1820. https://doi.org/10.3390/molecules26071820

Noda T, Isono N, Krivandin AV, Shatalova OV, Błaszczak W, Yuryev VP (2009) Origin of defects in assembled supramolecular structures of sweet potato starches with different amylopectin chain-length distribution. Carbohydr Polym 76(3):400–409. https://doi.org/10.1016/j.carbpol.2008.10.029

Obata N, Tabuchi H, Kurihara M, Yamamoto E, Shirasawa K, Monden Y (2022) Mapping of nematode resistance in hexaploid sweetpotato using a next-generation sequencing-based association study. Front Plant Sci 13. https://doi.org/10.3389/fpls.2022.858747

Okwuonu IC, Obidiegwu JE, Onyeka TJ, Egesi CN (2022) Genome engineering: a new frontier for improving toot and tuber crops. In: Obembe OO, Ekundayo EO, Okoli AS, Gidado A, Adetunji CO, Ibrahim AB, Ubi BE (eds) Agricultural biotechnology, biodiversity and bioresources conservation and utilization. CRC Press, Boca Raton

Oloka BM, da Silva PG, Amankwaah VA, Mollinari M, Pecota KV, Yada B, Olukolu BA, Zeng Z-B, Craig Yencho G (2021) Discovery of a major QTL for root-knot nematode (*Meloidogyne incognita*) resistance in cultivated sweetpotato (*Ipomoea batatas*). Theor Appl Genet 134(7): 1945–1955. https://doi.org/10.1007/s00122-021-03797-z

Palumbo F, Galvao AC, Nicoletto C, Sambo P, Barcaccia G (2019) Diversity analysis of sweet potato genetic resources using morphological and qualitative traits and molecular markers. Genes 10(11):E840. https://doi.org/10.3390/genes10110840

Panda V, Sonkamble M (2012) Phytochemical constituents and pharmacological activities of *Ipomoea batatas* L. (Lam) – a review. Int J Res Phytochem Pharmacol 2(1):25–34

Park S-C, Kim SH, Park S, Lee H-U, Lee JS, Park WS, Ahn M-J, Kim Y-H, Jeong JC, Lee H-S, Kwak S-S (2015) Enhanced accumulation of carotenoids in sweetpotato plants overexpressing *IbOr-Ins* gene in purple-fleshed sweetpotato cultivar. Plant Physiol Biochem 86:82–90. https://doi.org/10.1016/j.plaphy.2014.11.017

Park S, Kim HS, Jung YJ, Kim SH, Ji CY, Wang Z, Jeong JC, Lee H-S, Lee SY, Kwak S-S (2016) Orange protein has a role in phytoene synthase stabilization in sweetpotato. Sci Rep 6(1):33563. https://doi.org/10.1038/srep33563

Qin Z, Li A, Dong S, Wang Q, Hou F, Zhang H (2021) Comparative transcriptome analysis of hybrid population provides insights into starch content in sweet potato (*Ipomoea batatas* L.) storage root. Plant Mol Biol Report 39(4):673–684. https://doi.org/10.1007/s11105-021-01282-x

Rabbi IY, Udoh LI, Wolfe M, Parkes EY, Gedil MA, Dixon A, Ramu P, Jannink J-L, Kulakow P (2017) Genome-wide association mapping of correlated traits in cassava: dry matter and total carotenoid content. Plant Genome 10(3). https://doi.org/10.3835/plantgenome2016.09.0094

Rahman KAMM, Shukor Juraimi A, Rezwan Molla M, Asyraf Md Hatta M, Bin Sulaiman Z, Sultana S, Gaber A, Biswas B, Hossain A (2023) Molecular characterization of sweet potato (*Ipomoea batatas* [L.] Lam) germplasms for desirable traits by using simple sequence repeats markers. Biocell 47(1):227–237. https://doi.org/10.32604/biocell.2023.025135

Rodríguez-Villalón A, Gas E, Rodríguez-Concepción M (2009) Phytoene synthase activity controls the biosynthesis of carotenoids and the supply of their metabolic precursors in dark-grown Arabidopsis seedlings. Plant J 60(3):424–435. https://doi.org/10.1111/j.1365-313X.2009.03966.x

Rossel G, Kriegner A, Zhang DP (2001) From Latin America to Oceania: the historic dispersal of sweetpotato re-examined using AFLP. In: Scientist and farmer: partners in research for the 21st century – program report 1999–2000. International Potato Center, Lima, pp 315–321

Rosyara UR, De Jong WS, Douches DS, Endelman JB (2016) Software for genome-wide association studies in autopolyploids and its application to potato. Plant Genome 9(2). https://doi.org/10.3835/plantgenome2015.08.0073

Roullier C, Rossel G, Tay D, Mckey D, Lebot V (2011) Combining chloroplast and nuclear microsatellites to investigate origin and dispersal of New World sweet potato landraces. Mol Ecol 20(19):3963–3977. https://doi.org/10.1111/j.1365-294X.2011.05229.x

Roullier C, Duputié A, Wennekes P, Benoit L, Bringas VMF, Rossel G, Tay D, McKey D, Lebot V (2013) Disentangling the origins of cultivated sweet potato (*Ipomoea batatas* (L.) Lam.). PLoS One 8(5):e62707. https://doi.org/10.1371/journal.pone.0062707

Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, Lu Z, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Miller V, Mizrachi I, Ostell J, Panchenko A, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Slotta D, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Wang Y, John Wilbur W, Yaschenko E, Ye J (2010) Database resources of the National Center for biotechnology information. Nucleic Acids Res 38:D5–D16. https://doi.org/10.1093/nar/gkp967

Serang O, Mollinari M, Garcia AAF (2012) Efficient exact maximum a posteriori computation for Bayesian SNP genotyping in polyploids. PLoS One 7(2):e30906. https://doi.org/10.1371/journal.pone.0030906

Shirasawa K, Tanaka M, Takahata Y, Ma D, Cao Q, Liu Q, Zhai H, Kwak S-S, Cheol Jeong J, Yoon U-H, Lee H-U, Hirakawa H, Isobe S (2017) A high-density SNP genetic map consisting of a complete set of homologous groups in autohexaploid sweetpotato (*Ipomoea batatas*). Sci Rep 7(1):44207. https://doi.org/10.1038/srep44207

Spannagl M, Noubibou O, Haase D, Yang L, Gundlach H, Hindemitt T, Klee K, Haberer G, Schoof H, Mayer KFX (2007) MIPSPlantsDB – plant database resource for integrative and comparative plant genome research. Nucleic Acids Res 35:D834–D840. https://doi.org/10.1093/nar/gkl945

Spindel JE, Begum H, Akdemir D, Collard B, Redoña E, Jannink J-L, McCouch S (2016) Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. Heredity 116(4):395–408. https://doi.org/10.1038/hdy.2015.113

Steed LE, Truong V-D (2008) Anthocyanin content, antioxidant activity, and selected physical properties of flowable purple-fleshed sweetpotato purees. J Food Sci 73(5):S215–S221. https://doi.org/10.1111/j.1750-3841.2008.00774.x

Sun Q, Zybailov B, Majeran W, Friso G, Olinares PDB, van Wijk KJ (2009) PPDB, the plant proteomics database at Cornell. Nucleic Acids Res 37:D969–D974. https://doi.org/10.1093/nar/gkn654

The Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. Nature 475(7355):189–195. https://doi.org/10.1038/nature10158

Thompson PG, Hong LL, Ukoskit K, Zhu Z (1997) Genetic linkage of randomly amplified polymorphic DNA (RAPD) markers in sweetpotato. J Am Soc Hortic Sci 122(1):79–82. https://doi.org/10.21273/JASHS.122.1.79

Truong VD, Avula RY, Pecota KV, Yencho GC (2018) Sweetpotato production, processing, and nutritional quality. In: Siddiq M, Uebersax MA (eds) Handbook of vegetables and vegetable processing. Wiley, Chichester, pp 811–838

Ukoskit K, Thompson PG (1997) Autopolyploidy versus allopolyploidy and low-density randomly amplified polymorphic DNA linkage maps of sweetpotato. J Am Soc Hortic Sci 122(6):822–828. https://doi.org/10.21273/JASHS.122.6.822

Vargas PF, Engelking EW, Almeida LCFD, Ferreira EA, Charlo HCDO (2018) Genetic diversity among sweet potato crops cultivated by traditional farmers. Rev Caatinga 31(3):779–790. https://doi.org/10.1590/1983-21252018v31n329rc

Wadl PA, Olukolu BA, Branham SE, Jarret RL, Yencho GC, Jackson DM (2018) Genetic diversity and population structure of the USDA sweetpotato (*Ipomoea batatas*) germplasm collections using GBSpoly. Front Plant Sci 9. https://doi.org/10.3389/fpls.2018.01166

Wang A, Li R, Ren L, Gao X, Zhang Y, Ma Z, Ma D, Luo Y (2018) A comparative metabolomics study of flavonoids in sweet potato with different flesh colors (*Ipomoea batatas* (L.) Lam). Food Chem 260:124–134. https://doi.org/10.1016/j.foodchem.2018.03.125

Wang H, Wu Y, Zhang Y, Yang J, Fan W, Zhang H, Zhao S, Yuan L, Zhang P (2019) CRISPR/Cas9-based mutagenesis of starch biosynthetic genes in sweet potato (*Ipomoea batatas*) for the improvement of starch quality. Int J Mol Sci 20(19):4702. https://doi.org/10.3390/ijms20194702

Woolfe JA (1992) Sweet potato: an untapped food resource, 1st edn. Cambridge University Press, Cambridge.

Wu S, Lau KH, Cao Q, Hamilton JP, Sun H, Zhou C, Eserman L, Gemenet DC, Olukolu BA, Wang H, Crisovan E, Godden GT, Jiao C, Wang X, Kitavi M, Manrique-Carpintero N, Vaillancourt B, Wiegert-Rininger K, Yang X, Bao K, Schaff J, Kreuze J, Gruneberg W, Khan A, Ghislain M, Ma D, Jiang J, Mwanga ROM, Leebens-Mack J, Coin LJM, Yencho GC, Buell CR, Fei Z (2018) Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement. Nat Commun 9(1):4580. https://doi.org/10.1038/s41467-018-06983-8

Xiong J, Tang X, Wei M, Yu W (2022) Comparative full-length transcriptome analysis by Oxford Nanopore Technologies reveals genes involved in anthocyanin accumulation in storage roots of sweet potatoes (*Ipomoea batatas* L.). PeerJ 10:e13688. https://doi.org/10.7717/peerj.13688

Yada B, Brown-Guedira G, Alajo A, Ssemakula GN, Mwanga ROM, Yencho GC (2015) Simple Sequence Repeat marker analysis of genetic diversity among progeny of a biparental mapping population of sweetpotato. HortScience 50(8):1143–1147. https://doi.org/10.21273/HORTSCI.50.8.1143

Yada B, Brown-Guedira G, Alajo A, Ssemakula GN, Owusu-Mensah E, Carey EE, Mwanga ROM, Yencho GC (2017) Genetic analysis and association of simple sequence repeat markers with storage root yield, dry matter, starch and β-carotene content in sweetpotato. Breed Sci 67(2):140–150. https://doi.org/10.1270/jsbbs.16089

Yamamoto E, Shirasawa K, Kimura T, Monden Y, Tanaka M, Isobe S (2020) Genetic mapping in autohexaploid sweet potato with low-coverage NGS-based genotyping data. G3 Genes Genomes Genet 10(8):2661–2670. https://doi.org/10.1534/g3.120.401433

Yang J, Moeinzadeh MH, Kuhl H, Helmuth J, Xiao P, Haas S, Liu G, Zheng J, Sun Z, Fan W, Deng G, Wang H, Hu F, Zhao S, Fernie AR, Boerno S, Timmermann B, Zhang P, Vingron M (2017) Haplotype-resolved sweet potato genome traces back its hexaploidization history. Nat Plants 3(9):696–703. https://doi.org/10.1038/s41477-017-0002-z

Yang X, Yang S, Qi H, Wang T, Li H, Zhang Z (2020) PlaPPISite: a comprehensive resource for plant protein-protein interaction sites. BMC Plant Biol 20(1):61. https://doi.org/10.1186/s12870-020-2254-4

Zhang D, Cervantes J, Huamán Z, Carey E, Ghislain M (2000) Assessing genetic diversity of sweet potato (*Ipomoea batatas* (L.) Lam.) cultivars from tropical America using AFLP. Genet Resour Crop Evol 47(6):659–665. https://doi.org/10.1023/A:1026520507223

Zhang DP, Carbajulca D, Ojeda L, Rossel G, Milla S, Herrera C, Ghislain M (2001) Microsatellite analysis of genetic diversity in sweetpotato varieties from Latin America. In: Scientist and farmer: partners in research for the 21st century – program report 1999–2000. International Potato Center, Lima, pp 295–301

Zhang K, Wu Z, Li Y, Zhang H, Wang L, Zhou Q, Tang D, Fu Y, He F, Jiang Y, Yang H, Wang J (2014) ISSR-based molecular characterization of an elite germplasm collection of sweet potato (*Ipomoea batatas* L.) in China. J Integr Agric 13(11):2346–2361. https://doi.org/10.1016/S2095-3119(14)60779-6

Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, Banf M, Dai X, Martin GB, Giovannoni JJ, Zhao PX, Rhee SY, Fei Z (2016) iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. Mol Plant 9(12):1667–1670. https://doi.org/10.1016/j.molp.2016.09.014

Zhou Z, Tang J, Cao Q, Li Z, Ma D (2022) Differential response of physiology and metabolic response to drought stress in different sweetpotato cultivars. PLoS One 17(3):e0264847. https://doi.org/10.1371/journal.pone.0264847

Designing *Dioscorea* Genomes for Improved Nutritional and Pharmaceutical Properties

Ranjana Bhattacharjee

Contents

Abstract

Dioscorea species, commonly known as "Yams," belong to family *Dioscoreaceae* consisting of about 600 species distributed from Africa, Asia, the Caribbean's South America, and the South Pacific islands. The tuber of this genus is well known for their organoleptic properties, making them the most widely used food for carbohydrate, dietary supplements, and famine food. West Africa represents the region where yams are mostly consumed because of their underground and/or aerial tubers representing valuable sources of proteins, fats, and vitamins for millions of people. In addition to their nutritional properties, yams are the potential source of several bioactive compounds used in the treatment of many diseases, thus providing an opportunity for their use in pharmaceutical industries. However, the pharmaceutical properties of these bioactive compounds in yams need to be well researched and validated. Yams are known for the presence of secondary metabolites such as steroid saponin, allantoin, quinones, cyanidins, phenolics, and nitrogen-containing compounds as well as diosgenin representing different

R. Bhattacharjee (✉)
International Institute of Tropical Agriculture, Ibadan, Nigeria
e-mail: r.bhattacharjee@cgiar.org

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6_60

properties including antioxidant, hypoglycemic, hypolipidemic, antimicrobial, inflammatory, antiproliferative, androgenic, estrogenic, and contraceptive properties. This chapter summarizes the available information on nutritional and pharmaceutical properties of yams through published manuscripts and provides recommendations for future research directions for better utilization of yam tubers in human health and nutrition.

Keywords

Dioscorea · Yams · Traditional use · Nutritional properties · Pharmacological properties

1 Introduction

In the times of climate change and concerns about healthy diets across almost all the countries in the world, there is a need to develop sustainable food systems. The issues around food and nutritional security are of paramount importance today, and it has become necessary to diversify the present-day agricultural system as well as the food basket in this context. Root and tuber (RT) crops play a very important role after cereals in providing high calorific and superior carbohydrate contents as an alternate food source. Among these root and tuber crops, yams are well-known for providing both food and medicines to millions of people in the world, specifically in the tropical and subtropical regions in Africa, Asia, Latin America, the Caribbean, and the South Pacific islands. Yams represent the fourth most important root and tuber crop after potatoes, cassava, and sweet potatoes, contributing about 10% of total world's production of RT crops (Viruel et al. 2016) and second in West Africa after cassava (Lev and Shriver 1998). According to FAOSTAT 2021, the world's yam production was 75.14 million tonnes with West Africa, contributing up to 94% of production (71.22 million tonnes). In West Africa, three countries majorly contribute towards yam production, namely, Nigeria with 70.9% (50.4 million tonnes), Ghana with 11.7% (8.3 million tonnes), and Cote d'Ivoire with 11.0% (7.8 million tonnes) (FAOSTAT 2022) (Table 1). Yam production is growing at a rate of 4.07% in the world; however, this increase can be attributed to the increase in the area under yam cultivation (FAOSTAT 2022), as well as its recognition as food, income, and nutritional security crop. Additionally, yams play a very important role in the sociocultural environment of millions of people in West Africa (Obidiegwu and Akpabio 2017).

Yams (*Dioscorea* spp.) belong to the family Dioscoreaceae and subfamily Dioscoreoideae within the order Dioscoreales (Alexander and Coursey 1969), and the genus consists of about 600 species spread globally (Amanze et al. 2011). Majority of these species are known for their food, medicinal, and economic values, and play an important role as famine food or food system of millions of small and marginal rural families and forest-dwelling communities (Ngo Ngwe et al. 2015). The roots, tubers, and rhizomes of yams have been used as food and medicine since

Table 1 Production and area harvested for the first ten countries in the world for yams (FAOSTAT 2022)

Country	Area harvested (million Ha)	Production (million tonnes)
Nigeria	5.9	50.4
Ghana	0.45	8.3
Cote d'Ivoire	1.4	7.8
Benin	0.25	3.2
Togo	0.10	0.89
Cameroon	0.05	0.61
Central African Republic	0.06	0.48
Chad	0.05	0.46
Colombia	0.33	0.41
Papua New Guinea	0.21	0.37
Guinea	0.37	0.03
Africa	8.54	73.5
West Africa	8.2	71.2
World	8.7	75.1

prehistoric times by indigenous people (Singh 1960). Among all *Dioscorea* spp., only ten have been domesticated and are widely cultivated across Africa, Asia, and Latin America for food and income security (Scarcelli et al. 2017). The potential of yam as food can be attributed to its high levels of carbohydrates contributing about 200 dietary calories per person per day, while providing other nutritional components such as proteins, lipids, vitamins, and minerals (Lásztity et al. 1998). In addition, the pharmacological qualities of yams have also been reported due to the presence of bioactive compounds such as flavonoids, phenols, saponins, tannins, and alkaloids (Okwu and Ndu 2006), leading to antioxidant, immunomodulatory, estrogenic, angiotensin I-converting enzyme inhibiting, carbonic anhydrase and trypsin inhibiting, chitinase, anti-insect, anti-dust mite, lectin, and antiproliferative activities (Zhang et al. 2019). These compounds present in yams are believed to treat medicinal conditions such as inflammatory and cardiovascular diseases, aging disorders, menopause, cancers, and osteoporosis. With about 600 cultivated and wild species available in the genus *Dioscorea*, this crop has a great potential to provide starch and energy, as well as additional secondary bioactive compounds for the nutritional and medicinal needs of humans, mainly in rural and marginal areas. The high nutritional and pharmaceutical properties of this genus suggest that there is a need for conservation of all the species; however, currently the international yam collection maintained at the Genetic Resources Center, International Institute of Tropical Agriculture (IITA) represents only eight *Dioscorea* spp. One of the major challenges with conservation of such a large number of species is the widespread nature of this crop across the world and involvement of plant protection as well as exchange of genetic resources across borders. In addition, there is a need for more well-controlled research to validate the food, nutritional, and pharmacological value of different *Dioscorea* spp. spread across the world. This chapter summarizes the information

currently available on nutritional and pharmacological properties of different *Dioscorea* spp. while providing recommendations for research needs to generate appropriate and specific data on the importance of this crop.

2 Phytochemical Properties of Yams

Over the past several years, researchers across the world have evaluated and reported the phytochemical and nutritional qualities of different *Dioscorea* spp. It is obvious that the phytochemical abundance and associated nutritional qualities vary from one species to another and from one variety to another within each species, as well as the environmental conditions and agricultural practices under which the plants grow (Wu et al. 2015; Bekele and Bekele 2018) (Fig. 1). The most commonly available secondary metabolites in yams are saponins, and according to Ou-Yang et al. (2018),

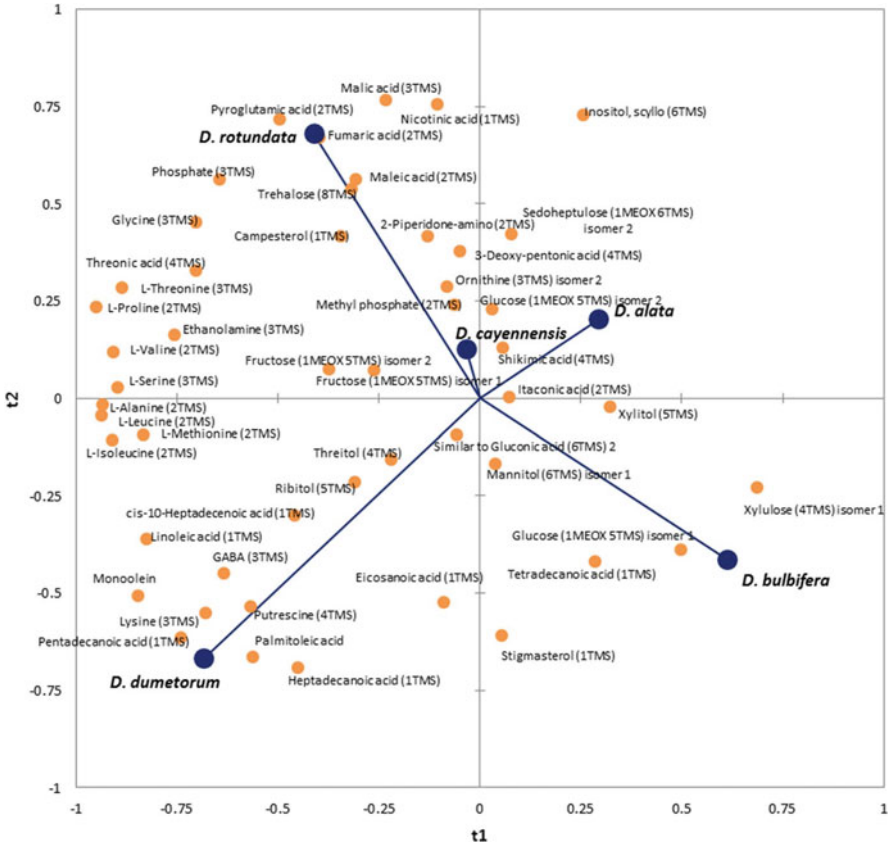


Fig. 1 The list of secondary metabolites present in different cultivated *Dioscorea* species (Price et al. 2017)

more than 100 steroidal saponins (such as stigmastanol, furostanol, spirostanol, cholestanol, ergostanol, and pregnanol glycosides) were isolated from various *Dioscorea* species. Additional compounds such as clerodane diterpenes, phenolics, cyanidins, quinones, diarylheptanoids, and nitrogen-containing compounds were also quantified (Ou-Yang et al. 2018). Similarly, Sautour et al. (2006) revealed over 50 saponins in 13 *Dioscorea* spp. including *D. cayenensis*, *D. bulbifera*, *D. colletti*, *D. futschauensis*, *D. deltoidea*, *D. panthaica*, *D. nipponica*, *D. pseudojaponica*, *D. parviflora*, *D. spongiosa*, *D. polygonoides*, *D. zingiberensis*, and *D. villosa*. Diosgenin (3-β-hydroxy-5-spirostene) is a primary saponin found in several plants, including *Dioscorea* species, and has several medicinal properties, such as hypolipidemic, antioxidant, anti-inflammatory, hypoglycemic, and antiproliferative activities (Jesus et al. 2016). The presence of diosgenin has been reported in several species of yams, making yams as one of the main raw materials in the pharmaceutical industry for production of anti-inflammatory, androgenic, estrogenic, and contraceptive steroidal drugs, used in the prevention/treatment of several diseases (Balandrin et al. 1985).

Although diosgenin has been reported in several yam species, its content varies from one species to another. Among all species, *D. barbasco* (Mexican wild yam) and *D. zingiberensis* (Chinese yam) are reported to consist high quantities of diosgenin (Yi et al. 2014; Kaimal and Kemper 1999), making both China and Mexico as world's highest (67%) diosgenin producers (Xiang et al. 2010). The presence of diosgenin has also been reported in *D. alata* (Jesus et al. 2016).

3 Nutritional Properties of Yams

Yams are considered to represent various dietary nutrients compared to other tuber crops, consisting of essential components such as starch, proteins, lipids, vitamins, minerals, and others (Arinathan et al. 2009; Mohan and Kalidass 2010) (Fig. 2). Among the researched *Dioscorea* spp., *D. alata* was reported to have high starch content (up to 84.3%) (Wanasundera and Ravindran 1994). Similarly, Afoakwa et al. (2013) evaluated the starch composition of seven cultivated yam species (*D. cayenensis*, *D. rotundata*, *D. alata*, *D. bulbifera*, *D. esculenta*, *D. prahensilis*, and *D. dumetorum*) and reported starch quantity ranging between 63.2% and 65.7%. In addition, yams contain dietary fiber which not only plays a vital role in the digestive system but also promotes physiological activities such as reduction in blood sugar and cholesterol levels, trapping of toxic substances, and increasing growth of microbial flora in the gut (Dhingra et al. 2012; Satija and Hu 2012; Slavin 2013; Sánchez-Zapata et al. 2015). Several dietary fibers such as hemicelluloses, cellulose, lignin, and pectins have been reported in yam (Abara et al. 2011; Shajeela et al. 2011).

Protein is one of the major nutritional components required for growth and development. In general, root and tuber crops consist of low proteins in comparison to legumes and cereals; however, among RT crops, yam represents higher dietary protein, although there is a wide variation in the protein content among different yam species (Chandrasekara and Josheph Kumar 2016). The essential amino acids present in yam tubers are phenylalanine and threonine (FAO 2022), while a recent

Nutrient	Potatoes	Cassava	Sweet potatoes	Yams
Proximate composition				
Energy (kJ)	322	670	360	494
Protein (g)	2	1.4	1.6	1.5
Total lipid (g)	0.09	0.28	0.05	0.17
Carbohydrates (g)	17	38	20	28
Total dietary fiber (g)	2.2	1.8	3	4.1
Sugar (g)	0.78	1.7	4.18	0.5
Minerals				
Calcium (mg)	12	16	30	17
Iron (mg)	0.78	0.27	0.61	0.54
Magnesium (mg)	23	21	25	21
Phosphorus (mg)	57	27	47	55
Potassium (mg)	421	271	337	816
Sodium (mg)	6	14	55	9
Vitamins				
Vitamin C (mg)	19.7	20.6	2.4	17.1
Thiamin (B1) (mg)	0.08	0.09	0.08	0.11
Vitamin E (mg)	0.01	0.19	0.26	0.39
Beta-carotene (μg)	1	8	8509	83
Fats				
Saturated fatty acids (g)	0.03	0.07	0.02	0.04
Monounsaturated fatty acids (g)	0	0.08	0	0.01
Polyunsaturated fatty acids(g)	0.04	0.05	0.01	0.08

Fig. 2 Comparison of nutritional properties of yams with other root and tuber crops (USDA 2015; Chandrasekara and Kumar 2016; Padhan and Panda 2020)

study on eight *Dioscorea* species reported the presence of aspartic acid and glutamic acid (Doss et al. 2019). Similarly, yams are known to contain essential minerals such as potassium which is higher than other RT crops such as cassava, potatoes, and sweet potatoes as well as cereals (maize, rice, and wheat) (Neela and Fanta 2019). Yams are also good source of other minerals such as calcium (Ezeocha et al. 2014), magnesium (Shajeela et al. 2011), iron (Otegbayo et al. 2018), manganese, zinc, and copper (Soto et al. 2014; Shajeela et al. 2011). There are reports that *Dioscorea* species consists of higher dietary fiber compared to potatoes, cassava, and sweet potatoes (Shanthakumari et al. 2008; Baah et al. 2009). The sugar and starch content of yams are reported to be less than potatoes and cassava (Baah et al. 2009; Afoakwa et al. 2013; Otegbayo et al. 2018), which is related to the presence of high dietary fiber, thus making it a suitable starchy crop for nutrition. Therefore, yams are reported to contribute 12% energy food after cassava (20%) (Otegbayo et al. 2018). The yam tubers are also rich in vitamin C (Udensi et al. 2008) and carotenoid content (Bhattacharjee et al. 2011).

4 Pharmacological Properties of Yams

There are not very many studies that explored the pharmacological properties of yams, although abundant information and traditional knowledge on the medicinal values of yams are prevalent in the tropical countries. Almost all *Dioscorea*

species contains high amounts of secondary metabolites that help in modulating metabolic processes in humans and animals, thus exhibiting beneficial health effects such as anti-oxidative, antihypertensive, anti-inflammatory, and anti-diabetic activities as well as inhibition of receptor activities and triggering enzyme activities (Correia et al. 2012; Price et al. 2017) (Table 2). A wide range of other compounds with medicinal properties as well as anti-nutritional properties are also present in yams such as phenolics, flavonoids, allantoin, dioscin, dioscorin, diosgenin, polyphenols, tannins, oxalate, saponin, and alkaloids have also been reported in yams by various researchers (Padhan et al. 2020; Price et al. 2016; Wadkar et al. 2014; Chiu et al. 2013; Mohan and Kalidass 2010; Shajeela et al. 2011; Kwon et al. 2015).

Saponins are glycosidic compounds with triterpenoid and steroidal properties, of which steroidal saponins are mainly present in different *Dioscorea* spp. (Escobar-Sánchez et al. 2016). Steroidal saponins can be further classified based on biological activities as spirostane, stigmastane, furostane, cholestane, ergostane, and pregnane families (Ali et al. 2013). Predominantly yams contain spirostane steroidal saponin known as diosgenin (Raju and Rao 2012; Cai et al. 2020), a precursor for the synthesis of steroidal drugs such as cortisone (Shah and Lele 2012), estrogen, and progesterone by the pharmaceutical companies (Balandrin et al. 1985; Cheng et al. 2015). In addition, diosgenin is used in the synthesis of hormonal drugs, including sex hormone, prevention of inflammation and malignant transformation, and regulation of T-cell immune responses as anticancer, neuro- and cardiovascular, as well as skin protection (Escobar-Sánchez et al. 2016; Raju and Rao 2012; Chen et al. 2015; Jesus et al. 2016; Tada et al. 2009). About 15 Dioscorea species are used as source of diosgenin with an estimated market value of USD 500 million (Price et al. 2017). However, its content varies considerably among different species and within each species, which may be attributed to the environmental conditions such as cultivation methods and storage conditions (Huai et al. 1989). In addition to diosgenin, several other steroidal saponins with pharmacological properties such as cytotoxic and antifungal properties have been reported in many *Dioscorea* species (Sautour et al. 2006).

Additionally, yams contain other bioactive compounds such as water-soluble polysaccharides and dioscorin (Myoda et al. 2006; Harijono et al. 2013). Dioscorin is the main storage protein of yam tubers classified as dioscorin A and dioscorin B that accounts for 90% of extractable water-soluble proteins (Hou et al. 2000; Conlan et al. 1995; Rachman 2011). Dioscorin acts as trypsin inhibitor, carbonic anhydrase, and antioxidant (Hou et al. 2001; Liu et al. 2007), and inhibits the activities of angiotensin-converting enzymes, thus controlling hypertension (Hsu et al. 2002) and activation of innate and adaptive immune system (Fu et al. 2006). There is also presence of alkaloids in yams, which are utilized in the pharmaceutical industry for their therapeutic activities including antimicrobial, antihypertensive, anticancer, anti-inflammatory, and several others. However, alkaloids also contribute to unwanted qualities such as bitterness and toxic substances causing nausea, dizziness, vomiting, and even trigger fatal paralysis (Banaag et al. 1997). In yams, the major alkaloid is dioscorine, a toxic molecule related to bitter taste and production of poisons (Broadbent and Schnieden 1958).

Table 2 Medicinal or pharmacological properties of few *Dioscorea* species (Obidiegwu et al. 2020)

Species name	Plant extract	Medicinal properties
D. alata	Tuber/bulb	Cure piles, gonorrhea, and leprosy; anti-inflammatory, purgative, diuretic, and antirheumatic properties; prevent cancer, reduce blood sugar, and diabetes
	Tuber	Antihelminthic properties
	Leaf	Fever
D. bartletti	Rhizome	Stagnation of blood and anemia
D. belophylla	Tuber	Treatment of fever, dysentery, headache, and malaria
D. bulbifera	Tuber	Treatment of dementia; treatment of diabetes, leprosy, tumors, microbial infections, and pig cysticercosis; antispasmodic, analgesic, aphrodisiac, diuretic, and rejuvenative tonic; effects on liver and heart; reduces carbuncles, lung abscesses, breast lumps, goiter, and abdominal pain; and oral contraceptive
	Raw tuber	Consumed as an appetizer, rheumatism, and aphrodisiac
	Aerial bulb	Oxidative stress-induced pathological disorders
D. belophylla	Tuber	Lowers blood cholesterol and reduces heart attack
D. cayenensis	Tuber	Antidiarrheal
D. collettii	Rhizome	Cervical carcinoma, urinary bladder carcinoma, and renal tumor
D. deltoidea	Tuber	Digestive disorders, sore throat, diarrhea, abdominal pains, wounds, burns, anemia, antirheumatic and treatment of ophthalmic conditions, birth control, and oral contraceptive
D. dumetorum	Tuber	Treatment of diabetes; control hyperlipidemia, hypercholesterolemia, and hyperketonemia; jaundice treatment
D. esculenta	Tuber	Inflammations, nervous disorders, respiratory infections, dysentery, and pain relief
D. hamiltonii	Tuber	Stomach ache and appetizer; management of diarrhea
D. hirtiflora	Tuber	Gonorrhea treatment
D. hispida	Leaf/root/ tuber	Treatment of mole, insect bites, and insomnia
	Tuber	Treatment of vomiting, indigestion, and serves as a purgative when consumed fresh; treatment of wounds and injuries; and ophthalmic ointment
D. japonica	Rhizome	Diarrhea and dysentery due to spleen deficiency, fatigue, wasting and thirsting, seminal emission, vaginal discharge and frequent urination, inflammation, asthma, rheumatoid arthritis, coughing, and wheezing
D. membranacea	Rhizome	Cancer
D. nipponica	Rhizome	Dissipation of lumps and goiter, clears heat, relieves toxicity, cools the blood, stops bleeding and coughing, calms sneezing, poisonous snake bites, bleeding due to blood heat and whooping cough, antirheumatic, analgesic, aids blood circulation, anti-diuretic, and aids digestion

(continued)

Table 2 (continued)

Species name	Plant extract	Medicinal properties
D. oppositifolia	Rhizomes/ tuber	Relief of menopausal syndromes and rejuvenation of early mothers
D. panthaica	Rhizome	Gastric diseases, bone injuries, rheumatic arthritis, and cardiovascular diseases
D. pentaphylla	Leaf/vine	Treatment of paralysis
	Tuber/flower/ young shoot	Rheumatism
	Tuber	Pain relief and reduce swelling, stomach disorders
D. polystachya	Rhizome	Consumptive cough and dysentery, aid for digestion and gastric motility, and for restraining nocturnal emissions
D. septemloba	Rhizome	Rheumatism, urethra, and renal infection
D. sylvatica	Tuber	Decoction used to treat cuts, wounds, and sores
D. trinervia	Tuber	Chronic diarrhea, asthma, and diabetes
D. villosa	Rhizome	Rheumatism, menstrual complaints, and perimenopausal symptoms
D. zingiberensis	Rhizome	Cough, anthrax, rheumatic heart disease, rheum, arthritis, tumefaction, and sprain

Water-soluble polysaccharides from yams are reported to reduce blood glucose (Fan et al. 2015) and cholesterol levels especially the LDL cholesterol (Harijono et al. 2013) while improving the immune system. The presence of flavonoids has also been reported in several species of yams (Padhan et al. 2020), which is known to exhibit antioxidant, antidiabetic, anti-inflammatory, antimicrobial, antidepressant, and antihypertensive properties (Sangeetha et al. 2016). The presence of active constituents in yams such as allantoin and dioscin is responsible for α-amylase and α-glucosidase activity providing antidiabetic, antioxidant, and anti-dyslipidemic properties (Niu et al. 2010). The presence of tannins in yams gives the bitter taste and is reported to treat diseases such as leukorrhea, rhinorrhea, healing of wound, and diarrhea (Eleazu et al. 2013).

In addition to medicinal properties, yam starch is being used extensively by the pharmaceutical industry for tablet and capsule formulation as well as thickening agent compared to potato starch (Zuluaga et al. 2007). Yam starch is also used as a binder in tablet formulations that contain both soluble and insoluble organic substances (Nasipuri 1979).

5 Anti-nutritional Properties of Yams

The presence of anti-nutritional factors influences the bioavailability of dietary nutrients such as proteins, minerals, and vitamins. Yam tubers are acrid in nature representing several anti-nutritional properties associated with skin irritation and inflammation of throat after consumption (Kumar et al. 2017). The presence of some bioactive compounds such as phenolic acids (Zhao et al. 2019), tannins, phytates,

and oxalates have been reported in different species of yams, which are referred to as anti-nutritional because of their toxic properties and bitterness associated with them (Poornima and Ravishankar Rai 2009). Similarly, the presence of trypsin inhibitors in yams (wild yam tubers have higher content than cultivated yams) inhibits the enzymatic activities of trypsin and chymotrypsin in the digestive tract, thus impairing protein digestion (Bhandari and Kawabata 2006; Padhan et al. 2020). Yams are identified as possible source of phenols and phenolic acids which are responsible for browning of tuber flesh when exposed to air. The presence of phenol and phenolic acids can decrease the digestibility of proteins, carbohydrates, and minerals (Padhan et al. 2018). Similarly, yam tubers are known to contain higher amount of phytates compared to other tuber crops which has a negative impact on the bioavailability of dietary minerals such as zinc, iron, calcium, copper, and magnesium (Polycarp et al. 2012; Padhan et al. 2018). Additionally, phytates also form insoluble complexes with other food components such as proteins, lipids, and carbohydrates, and inhibits the utilization of these nutrients (Kumar et al. 2010; Singh and Krikorian 1982). The presence of oxalates as insoluble calcium oxalate or soluble oxalate or combination of both have been reported in yams (Otegbayo et al. 2018), which may bind to minerals especially calcium, magnesium, and iron, resulting in their unavailability. Other effects could be linked with yam mucilage containing calcium oxalate that can cause severe skin irritation and throat inflammation (Noonan et al. 1999). It is reported that most of the anti-nutritional properties of yams are removed through washing, boiling, and processing (Otegbayo et al. 2018; Akin-Idowu et al. 2008).

6 Application of Multi-omics Tools for Nutritional and Pharmacological Properties of Yams

The genus *Dioscorea* with 600 species and several varieties within each species (at least for 10 cultivated species) provides a wide genetic diversity along with indigenous knowledge to explore the possibilities of application of modern omics tools for its improvement. Modern genomics tools, such as molecular markers linked to genes controlling traits such as nutritional traits, offer a golden opportunity to improve yam production (Mignouna et al. 2003) and application of in vitro micropropagation (Borges et al. 2004; Mantell et al. 1978; Ondo Ovono et al. 2007) and cleaning of viral disease (Wang et al. 2008; Ita et al. 2020), genetic transformation (Nyaboga et al. 2014), developing low cost, high throughput, scalable disease diagnostic tools (Filloux et al. 2018), and use of gene editing for trait improvement. However, the information on application of different omics technologies, including genomics, proteomics, etc., for yam improvement is limited (Sharma and Deswal 2016). Additionally, the yam breeding programs globally only target few cultivated *Dioscorea* species, mainly focusing on two species such as *D. rotundata* and *D. alata*. These programs target majorly traits, including tuber yield, resistance/tolerance to diseases, and quality-related traits, which may vary from one country to another (Mignouna et al. 2007). However, over the years, additional tuber

characteristics targeted to consumer preferences (tuber size and shape), higher dry matter, tuber flesh color, tuber flesh oxidation, and starch properties are being routinely measured for wider acceptance of newly developed varieties (De Koeyer et al. 2017; Darkwa et al. 2020). Some high-level traits requiring robust phenotyping such as resistance genes to viruses, anthracnose and nematodes, shrub-like or dwarf plant architecture, tuber bulking, tolerance to drought and heat, long shelf life of fresh tubers or processed food products, and culinary attributes suited to consumer needs for fresh and processed yam are not currently included in the breeding programs for developing new varieties.

Recently, there has been some progress on application of advanced molecular or omics tools to several species of *Dioscorea*. Currently based on short-reads, long-reads, and chromosome conformation sequencing, good-quality reference genomes have been released for four Dioscorea species such as *D. rotundata* (Tamiru et al. 2017), *D. alata* (Bredeson et al. 2022), *D. dumetorum* (Siadjeu et al. 2020), and *D. zingiberensis* (Cheng et al. 2021). This has enabled to develop high-quality linkage maps and identification of quantitative trait loci (QTLs) for several agronomic and quality traits such as sex determination systems (Tamiru et al. 2017; Cormier et al. 2019; Sugihara et al. 2020), anthracnose disease, tuber oxidative browning, tuber starch content (Bredeson et al. 2022), and postharvest hardening (Siadjeu et al. 2020). Similarly, Nakayasu et al. (2015) have performed comparative transcriptome analysis of high saponin-containing yams such as *D. esculenta* and *D. cayenensis*, and a low saponin-containing species, *D. alata*, to understand the biosynthesis of steroidal saponins. They identified the β-glucosidase (DeF26G1) gene for higher accumulation of saponins in *D. esculenta*. Wu et al. (2015) carried out transcriptome analysis of *D. alata* tubers (purple-flesh and white-flesh variety) and identified a total of 125,123 unigenes from cDNA libraries, out of which 61 unigenes encoded for various well-known enzymes in the flavonoid biosynthesis pathway. These genes were flavanone 3-hydroxylase (F3H), chalcone isomerase (CHS), dihydroflavonol 4-reductase (DFR), leucoanthocyanidin dioxygenase (LDOX), flavonoid 3′-monooxygenase (F3′H), and flavonol 3-O-glucosyl-transferase (UF3GT), which were more abundant in the purple-flesh variety than white flesh variety. The first report of genome-wide characterization was reported in *D. zingiberensis* by Zhou et al. (2018) and identified 4935 genes, 81 tRNAs, 661 miRNAs, and 69 rRNAs from them. Price et al. (2016) revealed 151 secondary metabolites from the polar extracts of the leaves of 19 *Dioscorea* species, including cultivated and wild species, using gas chromatography-mass spectrometry. This study provided insights about the presence of essential compounds such as shikimic acid (precursor utilized in the manufacture of the antiviral drug oseltamivir) in some species. Similarly, Price et al. (2017) carried out metabolite profiling of 49 accessions across 5 cultivated *Dioscorea* spp. (*D. alata*, *D. bulbifera*, *D. cayenensis*, *D. dumetorum*, and *D. rotundata*) and revealed about 200 bioactive compounds which can be used for chemo-typing of yams and integration with other omics studies for yam improvement. Again, Price et al. (2018) investigated the carotenoid profiling of the above accessions to determine the β-carotene content and provitamin A activity across species. This study clearly elucidated the lack of correlation

between yellow tuber flesh color and provitamin A content in yam as opposed to other crops like cassava. This study also profiled other isoprenoids, such as tocopherols and quinones as well as identified elite genotypes with high provitamin A content, and tentatively identified C25-epoxy-apocarotenoid persicaxanthin, which may influence tuber dormancy. Lebot et al. (2019) used high-performance thin-layer chromatography (HP-TLC) method for rapid quantification of individual sugars, allantoin, phenolic acids, catechins, and saponins in yam tuber flours. This technique was further used for rapid quantification of compounds related to tuber flour quality of eight *Dioscorea* species.

7 Genetic Engineering and Gene Editing in Yams

Genetic engineering and gene editing have emerged as important methodologies to improve crops including yam. The use of conventional breeding for improvement of several traits in yams is difficult, and application of transgenic methods provides an alternative and complimentary tool. Quain et al. (2011) developed a transient transformation for *D. rotundata* using agrobacterium, but this did not yield any transgenic plants. *Agrobacterium*-mediated gene delivery system is the most preferred method due to its ease of accessibility and ability to transfer low copies of DNA fragments carrying the desirable genes at higher efficiencies with minimal cost (Gelvin 2003; Shibata and Liu 2000). Nyaboga et al. (2014) developed an efficient and reproducible protocol for *Agrobacterium*-mediated transformation of *D. rotundata* which generated stable transgenic plants. This laid the foundation for successful implementation of genetic engineering and gene editing in yams, and the Genome-Enabled Platforms for Yam Project was launched in 2016 in collaboration between scientists at IITA and Iowa State University (https://www.nsf.gov/awardsearch/showA ward?AWD_ID = 1,543,888). Additionally, a genome-editing tool for yam using phytoene desaturase (a key enzyme in the β-carotene biosynthesis pathway, which converts the colorless phytoene to colored carotenoids) as a marker has been developed (Tripathi 2018; Syombua et al. 2021). Similarly, Feng et al. (2018) successfully applied clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9 (CRISPR/Cas9)-mediated targeted mutagenesis in *D. zingiberensis* using *Agrobacterium*-mediated transformation and aimed at the farnesyl pyrophosphate synthase gene (*Dzfps*) (an essential gene involved in the synthesis of secondary metabolites). They detected five types of mutations among the transformed plants and discovered that the transcript levels of *Dzfps* and the content of squalene in isolated mutants were drastically decreased. They concluded that CRISPR/Cas9 provided a rapid and efficient approach for targeted genome modification in *D. zingiberensis* which can be adopted to other *Dioscorea* species with slight modifications. The browning in yams due to the presence of polyphenol oxidase (PPO) is responsible for changes in the flavor, texture, and color, thus reducing the commercial value (Jia et al. 2015). The CRISPR/Cas system can be applied to generate heritable and stable mutations on the yam *PPO* loci without affecting other crop attributes. This technology for nutrition improvement has been

already proven through knockout of *PPO* gene in potatoes, mushrooms, and apples (Halterman et al. 2016; Nishitani et al. 2016; Waltz 2016) to create non-browning varieties. Similarly, raw yellow yam naturally has significantly low levels of beta carotene and thiamine compared to yellow cassava (Adepoju et al. 2018; Price et al. 2018), and the CRISPR/Cas approach could be applied to improve the nutritional potential of yam by redirecting the biosynthetic pathways to generate higher quantities of beneficial compounds and less anti-nutritional compounds (Sabzehzari et al. 2020). Lycopene cyclization during carotenoid biosynthesis involves two genes: *lycopene epsilon-cyclase* (*LCYE*) gene, which diverts the pathway towards biosynthesis of ε-carotenoids, and lycopene beta cyclase (LCYB), which catalyzes the formation of β-rings (Richaud et al. 2018). Thus, mutations on the yam *LCYE* gene could accumulate the flux of biosynthetic precursors towards the β branch and hence increase the β-carotene contents. Carotenoid biosynthetic pathway of banana has been manipulated by Kaur et al. (2020) using CRISPR/Cas9 to knock out the *LCYE* gene and obtained up to sixfold increase in the β-carotene contents. The thiamine content of yam could also be enhanced by CRISPR/Cas-based overexpression of the genes involved in the biosynthetic pathway, primarily *thi1*, *thi4*, and *thiC*. This has been already tested in Arabidopsis that led to simultaneous overexpression of *thi1*/*thi4* and *thiC* in the seed and leaf thiamine contents (Dong et al. 2015). Genome editing protocol can therefore be successfully incorporated in the yam breeding programs for targeted traits such as resistance to yam mosaic virus and anthracnose diseases, herbicide tolerance, and nematode resistance as well as quality traits in consultation with the breeders. It is, however, necessary to consider the ethical and regulatory implications related to genetically modified and gene-edited crops. The major challenge with CRISPR/Cas9 technology is that it may recognize sequences with up to five mismatched bases suggesting high rates of off-target effects (Roy et al. 2018). Some approaches such as DNA-RNA chimeric guides, Cpf1, a single RNA endonuclease that employs a T-rich PAM on the 5′ side of the guide, and specific point mutations have been developed to mitigate this limitation (Kleinstiver et al. 2016; Zetsche et al. 2016).

8 Conclusions and Recommendations

All the species (cultivated as well as wild) of *Dioscorea* genus make a significant contribution as a tuber crop to the diets of more than 300 million people living in the tropical countries. They are considered as an important source of energy food during famine. Despite their importance as a food source, yams contain several bioactive compounds including diosgenin which is useful for the treatment of various ailments and disorders as well as production of steroidal drugs. However, many of the pharmacological potential of yams needs to be validated through detailed investigations of about 600 cultivated and wild *Dioscorea* species occurring around the world. This will require a multidisciplinary and holistic approach including the involvement of indigenous natives who knows the properties of such a wide diversity of *Dioscorea* species, health specialists, and pharmacological experts

followed by conservation and sustainable use of the species diversity. Although several medicinal properties of yams such as hypolipidemic, hypoglycemic, antioxidant, anti-inflammatory, antimicrobial, antiproliferative, androgenic, estrogenic, and contraceptive have been reported, majority of these active constituents, however, have not been characterized. Similarly, the properties of all the secondary metabolites (alkaloids, saponin, flavonoids, tannins, and phenols) from different *Dioscorea* should be validated through standardized procedures to define the actual biological potentials of each species. It is essential to include clinical trials to understand the pharmacological properties of yams including the bioavailability of different bioactive compounds. Overall, this chapter suggests that the species of *Dioscorea* genus cannot be ignored and should be prioritized as a good alternative source of bioactive compounds that can prevent or alleviate both functional and infectious disease burden on pharmaceutical industries, which is a big challenge in developing countries. The lack of enough information on clinical trials with safety, toxicity, and efficacy on the use of yams in human health is a gap, and this will require concerted efforts among the researchers and donors to recognize yams as a valuable crop in the international market for its food, nutritional, medicinal, and pharmacological properties.

Although transgenics and modern genomic technologies have been applied in the last decade to complement classical breeding efforts in improving most crops, including vegetatively propagated crops, such research efforts, however, have been limited in yams. This slow progress can be attributed to the limitation of resources for yam research, lack of good genomic information, and genetic resources for yam. The recent sequencing of various yam genomes coupled with modern breeding tools offers unprecedented opportunities to accelerate yam biology research for accelerated genomics-assisted breeding. The genome editing technologies, in particular, offer possibilities for improving the yam nutritional composition, enhancing tolerance to abiotic stresses, and metabolic engineering of products applicable to pharmaceutical, biofuel, and agricultural industries. Majority of the traits in yams are controlled by minor genes with cumulative effect, and conventional QTL mapping and genome-wide association studies (GWAS) proves laborious and may not be validated. Therefore, the CRISPR/Cas technology coupled with the recent reference sequencing information of different *Dioscorea* species and pedigree analysis could facilitate rapid and efficient genetic characterization in yam.

Yams represent several inherent biological constraints, making it very arduous and lengthy endeavor for application of modern technologies, and significant milestones have been achieved in its improvement recently. A wide array of genomic resources, including markers for genetic studies, techniques for ploidy determination, and QTLs for several traits are currently available. Prospect for discovering additional molecular markers linked to genes and QTLs of additional target traits including nutritional and medicinal relevance, and applying marker-assisted selection in yam breeding is high due to reduction in sequencing cost and availability of reference genomes of four *Dioscorea* species. Recent advances and application of different "omics" tools such as genomics, transcriptomics, metabolomics, and phenomics will substantially increase the identification of functional gene (s) affecting important traits.

References

Abara AE, Tawo EN, Obi-Abang ME, Obochi GO (2011) Dietary fibre components of four common Nigerian *Dioscorea* species. Pak J Nutr 10:383–387

Adepoju OT, Boyejo O, Adeniji PO (2018) Effects of processing methods on nutrient and antinutrient composition of yellow yam (*Dioscorea cayenensis*) products. Food Chem 238: 160–165

Afoakwa EO, Polycarp D, Budu AS, Mensah-Brown H, Otoo E (2013) Variability in biochemical composition and cell wall constituents among seven varieties in Ghanian yam (*Dioscorea* sp.) germplasm. Afr J Food Agric Nutr Dev 13(4):8107–8127

Akin-Idowu PE, Odunola OA, Asiedu R, Maziya-Dixon B, Uwaifo AO (2008) Variation in nutrient and antinutrient contents of tubers from yellow yam (*Dioscorea cayenensis*) genotypes grown at two locations. J Food Agric Environ 6:95–100

Alexander J, Coursey DG (1969) The origins of yam cultivation. In: Ucko PJ, Dimbleby GH (eds) The domestication and exploitation of plants and animals. Proceedings of a meeting of the research seminar in archaeology and related subjects held at the Institute of Archaeology, London University. Gerald Duckworth, London

Ali Z, Smillie TJ, Khan IA (2013) Cholestane steroid glycosides from the root of *Dioscorea villosa* (wild yam). Carbohydr Res 370:86–91

Amanze NJ, Agbo NJ, Eke-Okoro ON, Njoku DN (2011) Selection of yam seeds from open pollination for adoption in yam (*Dioscorea rotundata* Poir) production zones in Nigeria. J Plant Breed Crop Sci 3(4):68–73

Arinathan V, Mohan VR, Maruthupandian A (2009) Nutritional and anti-nutritional attributes of some under–utilized tubers. Trop Subtrop Agroecosystems 10:273–278

Baah FD, Maziya-Dixon B, Asiedu R, Oduro I, Ellis WO (2009) Nutritional and biochemical composition of *D. alata* (*Dioscorea* spp.) tubers. J Food Agric Environ 7:373–378

Balandrin MF, Klocke JA, Wurtele ES, Bollinger WH (1985) Natural plant chemicals: sources of industrial and medicinal materials. Science 228:1154–1160

Banaag A, Honda H, Shono TE (1997) Effects of alkaloids from yam, *Dioscorea hispida* Schlussel, on feeding and development of larvae of the diamondback moth, *Plutella xylostella* (Lepidoptera: Yponomeutidae). Appl Entomol Zool 32:119–126

Bekele A, Bekele E (2018) Proximate and mineral composition variability in Ethiopian yam (*Dioscorea* spp.). J Food Nutr Sci 6:12–17

Bhandari MR, Kawabata J (2006) Cooking effects on oxalate, phytate, trypsin and a-amylase inhibitors of wild yam tubers of Nepal. J Food Compos Anal 19:524–530

Bhattacharjee R, Gedil M, Sartie A, Otoo E, Dumet D, Kikuno H, Kumar P, Asiedu R (2011) Dioscorea. In: Kole C (ed) Wild crop relatives: genomic and breeding resources, industrial crops. Springer, Berlin, pp 71–96

Borges M, Ceiro W, Meneses S, Aguilera N, Vázquez J, Infante Z, Fonseca M (2004) Regeneration and multiplication of *Dioscorea alata* germplasm maintained in vitro. Plant Cell Tissue Organ Cult 76(1):87–90

Bredeson JV, Lyons JB, Oniyinde IO, Okereke NR, Kolade O, Nnabue I, Nwadili CO, Hřibová E, Parker M, Nwogha J, Shu S, Carlson J, Kariba R, Muthemba S, Knop K, Barton GJ, Sherwood AV, Lopez-Montes A, Asiedu R, Jamnadass R, Muchugi A, Goodstein D, Egesi CN, Featherston J, Asfaw A, Simpson GG, Doležel J, Hendre PS, Deynze AV, Kumar PL, Obidiegwu JE, Bhattacharjee R, Rokhsar DS (2022) Chromosome evolution and the genetic basis of agronomically important traits in greater yam. Nat Commun 13:2001

Broadbent JL, Schnieden H (1958) A comparison of some pharmacological properties of dioscorine and dioscine. Br J Pharmacol Chemother 13:213–215

Cai B, Zhang Y, Wang Z, Xu D, Jia Y, Guan Y, Liao A, Liu G, Chun C, Li J (2020) Review article therapeutic potential of diosgenin and its major derivatives against neurological diseases: recent advances. Oxidative Med Cell Longev 2020:3153082

Chandrasekara A, Josheph Kumar T (2016) Roots and tuber crops as functional foods: a review on phytochemical constituents and their potential health benefits. Int J Food Sci 2016:3631647
Chen Y, Tang Y, Yu S, Han Y, Kou J, Liu B, Yu B (2015) Advances in the pharmacological activities and mechanisms of diosgenin. Chinese Journal of Natural Medicine 13:578–587
Cheng J, Chen J, Liu X, Li X, Zhang W, Dai Z, Lu L, Zhou X, Cai J, Zhang X, Jiang H, Ma Y (2021) The origin and evolution of the diosgenin biosynthetic pathway in yam. Plant Commun 2:100079
Cheng Y, Dong C, Huang C, Zhu Y (2015) Enhanced production of diosgenin from *Dioscorea zingiberensis* in mixed culture solid state fermentation with *Trichoderma reesei* and *Aspergillus fumigatus*. Biotechnol Biotechnol Equip 29:773–778
Chiu CS, Deng JS, Chang HY, Chen YC, Lee MM, Hou WC, Lee CY, Huang SS, Huang GJ (2013) Antioxidant and anti-inflammatory properties of Taiwanese yam (*Dioscorea japonica* Thunb. var. pseudojaponica (Hayata) Yamam.) and its reference compounds. Food Chem 141:1087–1096
Cormier F, Lawac F, Maledon E, Gravillion M-C, Nudol E, Mournet P, Vignes H, Chair H, Arnau G (2019) A reference high-density genetic map of greater yam (Dioscorea alata L.). Theoretical and Applied Genetics 132:1733–1744
Conlan RS, Griffiths LA, Napier JA, Shewry PR, Mantell S, Ainsworth C (1995) Isolation and characterisation of cDNA clones representing the genes encoding the major tuber storage protein (dioscorin) of yam (*Dioscorea cayenensis* Lam.). Plant Mol Biol 28:369–380
Correia RTP, Borges KC, Medeiros MF, Genovese MI (2012) Bioactive compounds and phenolic-linked functionality of powdered tropical fruit residues. Food Sci Technol Int 18:539–547
Darkwa K, Olasanmi B, Asiedu R, Asfaw A (2020) Review of empirical and emerging breeding methods and tools for yam (*Dioscorea* spp.) improvement: status and prospects. Plant Breed 139:474–497
De Koeyer D, Lopez-Montes A, Amele A, Bhattacharjee R, Abberton M, Maziya-Dixon B, Asiedu R (2017) AfricaYam – a project to strengthen yam breeding in West Africa. Retrieved from http://www.gcp21.org/wcrtc/abstracts/S13/S1304.DeKoeyerDavid.4214.3280.doc
Dhingra D, Michael M, Rajput H, Patil RT (2012) Dietary fibre in foods: a review. J Food Sci Technol 49:255–266
Dong W, Stockwell VO, Goyer A (2015) Enhancement of thiamin content in *Arabidopsis thaliana* by metabolic engineering. Plant Cell Physiol 56(12):2285–2296
Doss A, Tresina PS, Mohan VR (2019) Amino acid composition of wild yam (*Dioscorea* spp.). Food Res 3:617–621
Eleazu CO, Kolawole S, Awa E (2013) Phytochemical composition and antifungal actions of aqueous and ethanolic extracts of the peels of two yam varieties. Med Aromat Plants 2(4):128
Escobar-Sánchez ML, Sánchez-Sánchez L, Sandoval-Ramírez J (2016) Steroidal saponins and cell death in cancer. In: Ntuli T (ed) Cell death. InTechOpen, London, pp 331–351
Ezeocha VC, Nwogha JS, Ohuoba AN, Chukwu LI (2014) Evaluation of poultry manure application rates on the nutrient composition of *Dioscorea bulbifera* (aerial yam). Niger Food J 32:92–96
Fan Y, He Q, Luo A, Wang M, Luo A (2015) Characterization and anti-hyperglycemic activity of a polysaccharide from *Dioscorea opposita* Thunb roots. Int J Mol Sci 16(3):6391–6401
FAO (2022) Roots, tubers, plantains and bananas in human nutrition. Available online: http://www.fao.org/3/T0207E00.htm#Contents. Accessed on 1 June 2023
FAOSTAT – Food and Agriculture Organization of the United Nations (2022) Available online: http://www.fao.org/faostat/en/#data/QC. Accessed on 1 June 2023
Feng S, Song W, Fu R, Zhang H, Xu A, Li J (2018) Application of the CRISPR/Cas9 system in *Dioscorea zingiberensis*. Plant Cell Tissue Organ Cult 135(1):133–141
Filloux D, Fernandez E, Loire E, Claude L, Galzi S, Candresse T, Winter S, Jeeva ML, Makeshkumar T, Martin DP, Raumagnac P (2018) Nanopore-based detection and characterization of yam viruses. Sci Rep 8(1):1–11
Fu SL, Hsu YH, Lee PY, Hou WC, Hung LC, Lin CH, Chen CM, Huang YJ (2006) Dioscorin isolated from *Dioscorea alata* activates TLR4-signaling pathways and induces cytokine expression in macrophages. Biochem Biophy Res Commun 339:137–144

Gelvin SB (2003) Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool. Microbiol Mol Biol Rev 67:16–37

Halterman D, Guenthner J, Collinge S, Butler N, Douches D (2016) Biotech potatoes in the 21st century: 20 years since the first biotech potato. Retrieved from https://pubag.nal.usda.gov/catalog/4974488

Harijono TE, Apriliyanti MW, Afriliana A, Kusnadi J (2013) Physicochemical and bioactive characteristics of purple and yellow water yam (*Dioscorea alata*) tubers. Int J Pharmtech Res 5(4):1691–1701

Hou WC, Chen HJ, Lin YH (2000) Dioscorins from different *Dioscorea* species all exhibit both carbonic anhydrase and trypsin inhibitor activities. Bot Bull Acad Sin 41:191–196

Hou WC, Lee MH, Chen HJ, Liang WL, Han CH, Liu YW, Lin YH (2001) Antioxidant activities of dioscorin, the storage protein of yam (*Dioscorea batatas* Decne) tuber. J Agric Food Chem 49: 4956–4960

Hsu FL, Lin YH, Lee MH, Lin CL, Hou WC (2002) Both dioscorin, the tuber storage protein of yam (*Dioscorea alata* cv. Tainong no. 1), and its peptic hydrolysates exhibited angiotensin converting enzyme inhibitory activities. J Agric Food Chem 50:6109–6113

Huai ZP, Ding ZZ, He SA, Sheng CG (1989) Research on correlations between climatic factors and diosgenin content in *Dioscorea zingiberensis* wright. Yao Xue Xue Bao 24:702–706

Ita EE, Uyoh EA, Nakamura I, Ntui VO (2020) Efficient elimination of yam mosaic virus (YMV) from white yam (*Dioscorea rotundata* Poir.) by cryotherapy of axillary buds. S Afr J Bot 130: 123–129

Jesus M, Martins APJ, Gallardo E, Silvestre S (2016) Diosgenin: recent highlights on pharmacology and analytical methodology. J Anal Methods Chem 2016:4156293

Jia GL, Shi JY, Song ZH, Li FD (2015) Prevention of enzymatic browning of Chinese yam (*Dioscorea* spp.) using electrolyzed oxidizing water. J Food Sci 80(4):718–728

Kaimal A, Kemper KJ (1999) Wild yam (Dioscoreaceae). In: The Longwood Herbal Task Force and Center for holistic pediatric education and research. Syosset Public Library, Syossset

Kaur N, Alok AS, Kumar P, Kaur N, Awasthi P, Chaturvedi S, Pandey P, Pandey A, Pandey AK, Tiwari S (2020) CRISPR/Cas9 directed editing of lycopene epsilon-cyclase modulates metabolic flux for β-carotene biosynthesis in banana fruit. Metab Eng 59:76–86

Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zeng Z, Joung JK (2016) High-fidelity CRISPR–Cas9 nucleases with no detectable genome wide off-target effects. Nature 529: 490–495

Kumar S, Mahanti P, Rath SK, Patra JK (2017) Qualitative phytochemical analysis and antibacterial activity of *Dioscorea alata* L.: a nutraceutical tuber crops of rural Odisha. J Altern Med Res 3:122

Kumar V, Sinha AK, Makkar HPS, Becker K (2010) Dietary roles of phytate and phytase in human nutrition: a review. Food Chem 120:945–959

Kwon YK, Jie EY, Sartie A, Kim DJ, Liu JR, Min BW, Kim SW (2015) Rapid metabolic discrimination and prediction of dioscin content from African yam tubers using Fourier transform-infrared spectroscopy combined with multivariate analysis. Food Chem 166:389–396

Lásztity R, Hidvégi M, Bata Á (1998) Saponins in food. Food Rev Int 14:371–390

Lebot V, Malapa R, Molisal T (2019) Development of HP-TLC method for rapid quantification of sugars, catechins, phenolic acids and saponins to assess yam (*Dioscorea* spp.) tuber flour quality. Plant Genet Resour 17(1):62–72

Lev LS, Shriver AL (1998) A trend analysis of yam production, area, yield, and trade (1961–1996). In: Ligname, Plante Seculaire et Culture Davenir; Actes du Seminaire International CIRAD-INRA-ORSTOM-CORAF. CIRAD, Montpellier, pp 8–10

Liu YW, Shang HF, Wang CK, Hsu FL, Hou WC (2007) Immunomodulatory activity of dioscorin, the storage protein of yam (*Dioscorea alata* cv. Tainong no. 1) tuber. Food Chem Toxicol 45: 2312–2318

Mantell SH, Haque SQ, Whitehall AP (1978) Clonal multiplication of *Dioscorea alata* L. and *Dioscorea rotundata* Poir. yams by tissue culture. J Hortic Sci 53(2):95–98

Mignouna HD, Abang MM, Fagbemi SA (2003) A comparative assessment of molecular marker assays (AFLP, RAPD and SSR) for white yam (*Dioscorea rotundata*) germplasm characterization. Ann Appl Biol 142(3):269–276

Mignouna HD, Abang MM, Asiedu R (2007) Advances in yam (*Dioscorea* spp.) genetics and genomics. In: Proceedings of the 13th ISTRC symposium, Arusha, pp 72–81

Mohan VR, Kalidass C (2010) Nutritional and antinutritional evaluation of some unconventional wild edible plants. Trop Subtrop Agroecosystems 12:495–506

Myoda T, Matsuda Y, Suzuki T, Nakagawa T, Nagai T, Nagashima T (2006) Identification of soluble proteins and interaction with mannanin mucilage of *Dioscorea* opposite Thunb. (Chinese yam tuber). Food Sci Technol Res 12(4):299–302

Nakayasu M, Kawasaki T, Lee H, Sugimoto Y, Onjo M, Muranaka T, Mizutani M (2015) Identification of furostanol glycoside26-O-β-glucosidase involved in steroidal saponin biosynthesis from *Dioscorea esculenta*. Plant Biotechnol 32:1015–1023

Nasipuri R (1979) Evaluation of yam starch as tablet binder and disintegrant. Niger J Pharm 10: 182–187

Neela S, Fanta SW (2019) Review on nutritional composition of orange – fleshed sweet potato and its role in management of vitamin A deficiency. Food Sci Nutr 7:1920–1945

Ngo Ngwe MFS, Omokolo DN, Joly S (2015) Evolution and phylogenetic diversity of yam species (*Dioscorea* spp.): implication for conservation and agricultural practices. PLoS One 10(12):1–13

Nishitani C, Hirai N, Komori S, Wada M, Okada K, Osakabe K, Yamamoto T, Osakabe Y (2016) Efficient genome editing in apple using a CRISPR/Cas9 system. Sci Rep 6(1):31481

Niu CS, Chen W, Wu HT, Cheng KC, Wen YJ, Lin KC, Cheng JT (2010) Decrease of plasma glucose by allantoin, an active principle of yam (*Dioscorea* spp.) in spteptozotocin-induced diabetics rats. J Agric Food Chem 22:12031–12035

Noonan SC, Hons GPS, Nutr NZR (1999) Oxalate content of foods and its effect on humans. Asia Pac J Clin Nutr 8:64–74

Nyaboga E, Tripathi JN, Manoharan R, Tripathi L (2014) Agrobacterium-mediated genetic transformation of yam (*Dioscorea rotundata*): an important tool for functional study of genes and crop improvement. Front Plant Sci 5:1–14

Obidiegwu JE, Akpabio EM (2017) The geography of yam cultivation in southern Nigeria: exploring its social meanings and cultural functions. J Ethn Foods 4:28–35

Obidiegwu JE, Lyons JB, Chilaka CA (2020) The *Dioscorea* genus (yams): an appraisal of nutritional and therapeutic potentials. Foods 9:1304

Okwu DE, Ndu CU (2006) Evaluation of the phytonutrients, mineral and vitamin contents of some varieties of yam. Int J Mol Med Adv Sci 2:199–203

Ondo Ovono P, Kevers C, Dommes J (2007) Axillary proliferation and tuberisation of *Dioscorea cayenensis-D. rotundata* complex. Plant Cell Tissue Organ Cult 91(2):107–114

Otegbayo BO, Oguniyan DJ, Olunlade BA, Oroniran OO, Atobatele OE (2018) Characterizing genotypic variation in biochemical composition, anti-nutritional and mineral bioavailability of some Nigerian yam (*Dioscorea* spp.) landraces. J Food Sci Technol 55:205–216

Ou-Yang SH, Jiang T, Zhu L, Yi T (2018) *Dioscorea nipponica* Makino: a systematic review on its ethnobotany, phytochemical and pharmacological profiles. Chemical Cent J 12:1–18

Padhan B, Panda D (2020) Potential of neglected and under-utilized yams (*Dioscorea* spp.) for improving nutritional security and health benefits. Front Pharmacol 11:496

Padhan B, Biswas M, Dhal NK, Panda D (2018) Evaluation of mineral bioavailability and heavy metal content in indigenous food plant wild yams (*Dioscorea* spp.) from Koraput, India. J Food Sci Technol 55(11):4681–4686

Padhan B, Biswas M, Panda D (2020) Nutritional, anti-nutritional and physico-functional properties of wild edible yam (*Dioscorea* spp.) tubers from Koraput, India. Food Biosci 34:100527

Polycarp D, Afoakwa EO, Budu AS, Otoo E (2012) Characterization of chemical composition and anti-nutritional factors in seven species within the Ghanaian yam (*Dioscorea*) germplasm. Int Food Res J 19(3):985–992

Poornima GN, Ravishankar RV (2009) Evaluation of phytonutrients and vitamin contents in a wild yam, *Dioscorea belophylla* (Prain) Haines. Afr J Biotechnol 8:971–973

Price EJ, Wilkin P, Sarasan V, Fraser PD (2016) Metabolite profiling of *Dioscorea* (yam) species reveals underutilised biodiversity and renewable sources for high-value compounds. Sci Rep 6:29136

Price EJ, Bhattacharjee R, Lopez-Montes A, Fraser PD (2017) Metabolite profiling of yam (*Dioscorea* spp.) accessions for use in crop improvement programmes. Metabolomics 13:144

Price EJ, Bhattacharjee R, Lopez-Montes A, Fraser PD (2018) Carotenoid profiling of yams: clarity, comparisons and diversity. Food Chem 259:130–138

Quain MD, Egnin M, Bey B, Thompson R, Bonsi C (2011) Transgenic potential of *Dioscorea rotundata*, using agrobacterium-mediated genetic transformation. Asp Appl Biol 110:71–79

Rachman F (2011) Antihypertensive effects of water-soluble polysaccharide binding dioscorin from Gadung (*Dioscorea hispida* Dennst.). Bachelor thesis Malang, Indonesia: food science and technology, agriculture technology faculty, Brawijaya University, pp 1–3

Raju J, Rao CV (2012) Diosgenin, a steroid saponin constituent of yams and fenugreek: emerging evidence for applications in medicine. In: Rasooli I (ed) Bioactive compounds in phytomedicine. IntechOpen, London, pp 125–142

Richaud D, Stange C, Gadaleta A, Colasuonno P, Parada R, Schwember AR (2018) Identification of lycopene epsilon cyclase (LCYE) gene mutants to potentially increase β-carotene content in durum wheat (*Triticum turgidum* L. ssp. Durum) through TILLING. PLoS One 13(12):e0208948

Roy B, Zhao J, Yang C, Luo W, Xiong T, Li Y, Fang X, Gao G, Singh CO, Madsen L, Zhou Y, Kristiansen K (2018) CRISPR/Cascade 9-mediated genome editing-challenges and opportunities. Front Genet 9(240):1–12

Sabzehzari M, Zeinali M, Naghavi MR (2020) CRISPR-based metabolic editing: next-generation metabolic engineering in plants. Gene 759:144993

Sánchez-Zapata E, Viuda-Martos M, Fernández-López J, Pérez-Alvarez JA (2015) Resistant starch as functional ingredient. In: Ramawat KG, Mérillon JM (eds) Polysaccharides. Springer International Publishing, Cham, pp 1911–1931

Sangeetha SKS, Umamaheswari S, Reddy MU, Narayana KS (2016) Flavonoids: therapeutic potential of natural pharmacological agents. Int J Pharm Sci Res 7:3924–3930

Satija A, Hu FB (2012) Cardiovascular benefits of dietary fiber. Curr Atheroscler Rep 14:505–514

Sautour M, Miyamoto T, Lacaille-Dubois MA (2006) Steroidal saponins and fla-van-3-ol glycosides from *Dioscorea villosa*. Biochem Syst Ecol 34:60–63

Scarcelli N, Chaïr H, Causse S, Vesta R, Couvreur TL, Vigouroux Y (2017) Crop wild relative conservation: wild yams are not that wild. Biol Conserv 210:325–333

Shah H, Lele S (2012) Extraction of diosgenin, a bioactive compound from natural source *Dioscorea alata* var purpurea. J Anal Bioanal Tech 3(4):1–3

Shajeela PS, Mohan VR, Jesudas LL, Soris PT (2011) Nutritional and antinutritional evaluation of wild yam (*Dioscorea* spp.). Trop Subtrop Agroecosystems 14:723–730

Shanthakumari S, Mohan VR, de Britto J (2008) Nutritional evaluation and elimination of toxic principles in wild yam (*Dioscorea* spp.). Trop Subtrop Agroecosystems 8:319–325

Sharma S, Deswal R (2016) Genomic and proteomic tools for understanding mysterious protein Dioscorin from *Dioscorea* tuber. In: Zargar SM, Rai V (eds) Plant omics and crop breeding. Apple Academic Press, Palm Bay, pp 97–114

Shibata D, Liu YG (2000) Agrobacterium-mediated plant transformation with large DNA fragments. Trends Plant Sci 5:354–357

Siadjeu C, Pucker B, Viehöver P, Albach DC, Weisshaar B (2020) High contiguity de novo genome sequence assembly of trifoliate yam (*Dioscorea dumetorum*) using long read sequencing. Genes 11:274

Singh M, Krikorian AD (1982) Inhibition of trypsin activity in vitro by phytate. J Agric Food Chem 30:799–800

Singh SC (1960) Some wild plants of food value in Nepal. J Tribhuban Univ Kathmandu 4:50–56

Slavin J (2013) Fiber and prebiotics: mechanisms and health benefits. Nutrients 5:1417–1435

Soto LJM, González JV, Nicanor AB, Cruz LG, Fernández JY (2014) Chemical characterisation and nutritional evaluation of mountain yam (*Dioscorea remotiflora* Kunth) tubers. Adv Biores 5:153–160

Sugihara Y, Darkwa K, Yaegashi H, Natsume S, Shimizu M, Abe A, Hirabuchi A, Ito K, Oikawa K, Tamiru-Oli M, Ohta A, Matsumoto R, Agre P, DeKoyer D, Pachakkil B, Yamanaka S, Muranaka S, Takagi H, White B, Asiedu R, Innan H, Asfaw A, Adebola P, Terauchi R (2020) Genome analyses reveal the hybrid origin of the staple crop white Guinea yam (Dioscorea rotundata). Proceeding of National Academy of Sciences USA 117(50):31987–31989

Syombua ED, Zhang Z, Tripathi JN, Ntui VO, Kang M, George OO, Tripathi L (2021) A CRISPR/Cas9-based genome-editing system for yam (*Dioscorea* spp.). Plant Biotechnol J 19:645–647

Tada Y, Kanda N, Haratake A, Tobiishi M, Uchiwa H, Watanabe S (2009) Novel effects of diosgenin on skin aging. Steroids 74:504–511

Tamiru M, Natsume S, Takagi H, White B, Yaegashi H, Shimizu M, Yoshida K, Uemura A, Oikawa K, Abe A, Urasaki N, Matsumura H, Babil P, Yamanaka S, Matsumoto R, Muranaka S, Girma G, Lopez-Montes A, Gedil M, Bhattacharjee R, Abberton M, Lava Kumar P, Rabbi I, Tsujimura M, Terauchi R (2017) Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination. BMC Biol 15(1):86

Tripathi L (2018) Prospects of genetic engineering and gene editing in yam improvement. Paper presented at the 17th plant and animal genome conference, San Diego, USA, 15th January 2018

Udensi EA, Oselebe HO, Iweala OO (2008) Investigation of chemical composition and functional properties of water yam (*Dioscorea alata*): effect of varietal differences. Pak J Nutr 7:342–344

USDA (2015) Dietary guidelines advisory committee. Scientific report of the guidelines advisory committee. USDA and US Department of Health and Human Services, Washington, DC. https://www.dietaryguidelines.gov/sites/default/files/2019-05/2015-2020_Dietary_Guidelines.pdf. Accessed 8 June 2023

Viruel J, Segarra-Moragues JG, Raz L, Forest F, Sanmartin I, Catalan P (2016) Late cretaceous-early eocene origin of yams (*Dioscorea*, Dioscoreaceae) in the Laurasian Palaearctic and their subsequent Oligocene-Miocene diversification. J Biogeogr 43(4):750–762

Wadkar KA, Magdum CS, Patil SS, Naikwade NS (2014) Anti-diabetic potential and Indian medicinal plants. J Herb Med Toxicol 2:45–50

Waltz E (2016) Gene-edited CRISPR mushroom escapes US regulation. Nature News 532(7599):293

Wanasundera JPD, Ravindran G (1994) Nutritional assessment of yam (*Dioscorea alata*) tubers. Plant Foods Hum Nutr 46:33–39

Wang Q, Cuellar WJ, Rajamäki ML, Hirata Y, Valkonen JPT (2008) Combined thermotherapy and cryotherapy for efficient virus eradication: relation of virus distribution, subcellular changes, cell survival and viral RNA degradation in shoot tips. Mol Plant Pathol 9(2):237–250

Wu ZG, Jiang W, Mantri N, Bao XQ, Chen SL, Tao ZM (2015) Transciptome analysis reveals flavonoid biosynthesis regulation and simple sequence repeats in yam (*Dioscorea Alata* L.) tubers. BMC Genomics 16:1–12

Xiang L, JianZhong M, YunDong S (2010) Research progress and prospects of *Dioscorea* and diosgenin. Chem Ind For Prod 30:107–112

Yi T, Fan LL, Chen HL, Zhu GY, Suen HM, Tang YN, Zhu L, Chu C, Zhao ZZ, Chen HB (2014) Comparative analysis of diosgenin in *Dioscorea* species and related medicinal plants by UPLC-DAD-MS. BMC Biochem 15:9

Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Vols SE, Joung J, van der Oost J, Regev A, Koonin EV, Zhang F (2016) Cpf1 is a single-RNA-guided endonuclease of a class 2 CRISPR-Cas system. Transgenic Res 25:207

Zhang L, Ng TB, Lam JKW, Wang SW, Lao L, Zhang KYZ, Zhang SCW (2019) Research and development of proteins and peptides with therapeutic potential from yam tubers. Curr Protein Pept Sci 20:277–284

Zhao C, Zhou B, Miao J, Li X, Jing S, Zhang D, Wang JY, Li X, Huang L, Gao W (2019) Multicomponent analysis and activities for evaluation of *Dioscorea oppositifolia* and *Dioscorea hamiltonii*. Food Agric Immunol 30:1148–1161

Zhou W, Li B, Li L, Ma W, Liu Y, Feng S, Wang Z (2018) Genome survey sequencing of *Dioscorea Zingiberensis*. Genome 61:567–574

Zuluaga MF, Baena Y, Mora CE, D'León LFP (2007) Physicochemical characterization and application of yam (*Dioscorea cayenensis-rotundata*) starch as a pharmaceutical excipient. Starch-Starke 59:307–317

Index

C. Kole (ed.), *Compendium of Crop Genome Designing for Nutraceuticals*,
https://doi.org/10.1007/978-981-19-4169-6

N