Livestock Diseases and Management
Series Editor: Yashpal Singh Malik

Chandra Sekhar Mukhopadhyay
Ratan Kumar Choudhary
Harsh Panwar
Yashpal Singh Malik *Editors*

Biotechnological Interventions Augmenting Livestock Health and Production

Springer

Livestock Diseases and Management

This book series discusses the various infectious diseases affecting the livestock, principle of the disease control, and specific disease management. It discusses the existing strategies to control infectious disease includes animal management programs, vaccination, targeted antimicrobial use, and food hygiene.

Despite public health and veterinary public health improvement within the last century, animal populations remain vulnerable to health threats caused by infectious diseases.

It reviews the current understanding of the zoonotic, emerging, and transboundary animal infections in relation to their transmission, epidemiology, clinical and pathological effects, diagnosis and treatment.

It also examines the recent advancements in the veterinary diagnostics including the existing capabilities, constraints, opportunities, and future potentials. In addition, it elaborates on the conventional and recombinant vaccines that are used in the veterinary medicines and the molecular approaches that have led to the development of new vaccines in recent years. A volume focusing on the various water- and food-borne diseases and its impact on the domestic animals is also a part of this series.

The book series examines the emergence of antimicrobial resistance in livestock, ongoing global surveillance, and monitoring program, its impact on the animal-human interface and strategies for combating resistance.

Chandra Sekhar Mukhopadhyay •
Ratan Kumar Choudhary • Harsh Panwar •
Yashpal Singh Malik
Editors

Biotechnological Interventions Augmenting Livestock Health and Production

Springer

Editors
Chandra Sekhar Mukhopadhyay
College of Animal Biotechnology
Guru Angad Dev Veterinary and Animal Sciences University
Ludhiana, Punjab, India

Ratan Kumar Choudhary
College of Animal Biotechnology
Guru Angad Dev Veterinary and Animal Sciences University
Ludhiana, Punjab, India

Harsh Panwar
College of Dairy Science and Technology
Guru Angad Dev Veterinary and Animal Sciences University
Ludhiana, Punjab, India

Yashpal Singh Malik
College of Animal Biotechnology
Guru Angad Dev Veterinary and Animal Sciences University
Ludhiana, Punjab, India

ISSN 2662-4346 ISSN 2662-4354 (electronic)
Livestock Diseases and Management
ISBN 978-981-99-2208-6 ISBN 978-981-99-2209-3 (eBook)
https://doi.org/10.1007/978-981-99-2209-3

The opinion expressed in this compilation is entirely by the respective authors and in any case, does not reflect the views of the editors of this publication

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

The book is dedicated to the students, researchers, and professionals in the fields of biotechnology and molecular biology.

Foreword

Advances in biotechnology have catered to wide areas of science, namely, food science, medicine, agriculture, environmental and animal science. The book *Biotechnological Augmenting Livestock Health and Production* is a nice compilation of advancements in biotechnological interventions in the field of animal science aimed at improving health and productivity. Editors have great knowledge and experience in putting together updated information on animal biotechnology through eminent experts.

This advanced-level book is well organized into three sections, viz. Functional Genomics and Proteomics, Microbial Genomics and Novel Biotechnological approach to enhance animal health, and Bioinformatics, Big Data, and Integrated Omics. The genetic makeup of the animal with functional phenotypes determines how the traits—either production or overall health, are going to be present in the herd. Disease prevention is a vital tool in animals' survival, and hence better health, longer productive life, and ultimately better food production. Another novel area of biotechnology is a stem cell and genome editing techniques. At last, a separate section on bioinformatics and multi-omics approach to a deep understanding of system biology with artificial intelligence learning open new avenues for understating livestock's system biology. Precision livestock and sustainable animal farming will be more realized upon the knowledge of system biology.

It is good to see that the editors from different areas of expertise have joined hands together to publish the book. Dr. Yashpal Singh Malik, for example, has contributed to the field of animal health, apart from other contributions in publishing dozens of books and hundreds of scientific publications. Moreover, the diverse fields of editors' expertise made this book a reference material for animal scientists, veterinarians, and research professionals, including students. The book may provide insights to young researchers in applying new interventions of biotechnology to enhance the health and production potential of livestock. The existence of diseases is inevitable, but finding the solution with newer technology will pave the way toward finding the solution.

Finally, I have been fortunate enough to know many novels and emerging axes of biotechnology that could help solve upcoming problems. Augmenting livestock health and productivity can be achieved by (1) creating disease-resistant animals using genome editing techniques, (2) treating deadly diseases of animals by regenerative medicine, and (3) preventing animals from getting infections using next-generation vaccines.

I convey my best wishes to the editors, authors, and contributors of this excellent compilation.

Preface

Scientific research is one of the most exciting and rewarding occupations. It is like a voyage of discovery into unknown lands, seeking not new territory but new knowledge. It should appeal to those with a good sense of adventure—Frederick Sanger

Genomics has played a critical role in addressing various problems in medical, veterinary, and agricultural sciences. The era of molecular markers (the 1980s onwards) has enabled humankind to conceptualize and fructify the much-coveted human genome project. This book focuses on various advanced perspectives of biotechnology in different areas of animal welfare which is intricately woven with the well-being of humankind. Chapters are divided into three subsections: (i) Functional genomics and proteomics (ii) Microbial genomics and novel biotechnological approaches to enhance animal health; and (iii) Bioinformatics, big data, and integrated omics. The first section focuses on various cutting-edge applications of functional genomics techniques and proteomics in animal sciences and microbial biotechnology. The first chapter "Genome editing: applications in animal health" elaborates on the importance of genome editing tools in understanding diseases that produce disease-resistant livestock animals. Mathapati and Singh, the authors of the chapter have elaborated on the applications of modern biotechnology tools of "genome editing" in preserving animal health through various means like selective breeding, development diagnostics, therapeutics, and other disease prevention modalities like effective vaccines. The second chapter on stem cell therapy, written by Saini and team, discusses the future potential and current applications of stem cell therapy to treat animal diseases. The next chapter confers information about proteomics and mechanical regulation of the onset of a disease like mastitis. The fourth chapter, contributed by Dr. Deepali Kalambhe and team, has contributed to an important aspect: antimicrobial susceptibility testing. In general, the chapter covers an overview of the molecular mechanism underlying microbial resistance against antimicrobials and various advanced methods for testing antimicrobial susceptibility. Dr. Wenli Li has contributed a very interesting concept on the impact of copy number variation (CNV), spanned throughout the genome, toward genomic selection. The sixth chapter elaborates on the concept of genomic selection and the

possible roadmap to implement it. The last chapter of the first section discusses the next-generation sequence-based biomarkers being used in the field of animal sciences.

The second section of this book "Microbial genomics and novel biotechnological approaches to enhance animal health" enlightens on microbial genomics and novel biotechnological tools for improving animal health. This section is composed of seven chapters having information on microbial diversity in animal health, how to measure microbial diversity, and the modulation of microbial genomics to improve animal health and mitigate carbon footprints by reducing methane production by livestock animals. In Chap. 8, the group led by Dholpuria discusses advanced genomics innovations in understanding microbial diversity, antibiotic resistance, and host–microbe interactions. Chapter 9 by Sangwan and coworkers describes different basic and advanced methods for measuring microbial diversity. The next chapter provides an understanding of the scope of microbial diversity and its relevance to animal health. The chapter also discusses symbiotic relationship management, microbiome manipulation, and customized microbial-based therapies. Chapter 11, written by Jain and colleagues, focuses on the understanding of relatedness between microbiota, animal gut, and brain, the impact of microbial invasion on diseases involving the gut-brain axis in animals and humans, and reviews some of the associated diseases with the gut-brain axis caused by microorganisms. Chapter 12 has been contributed by Kalambhe and coworkers, and it presents different genome-based bacterial antibiotic resistance prediction strategies. Debbarma and team have provided an update on microbial genomics and modulation in ruminants and how the reviewed methods could help prevent methane emissions in Chap. 13. The penultimate chapter of the second section is quite interesting as it portrays the role of vaccines in combating the COVID-19 pandemic. The authors Gautam and coworkers have elaborated on how the new generation vaccines can be generated in a very limited time frame to tackle future pandemics to save humanity and also for animal welfare. The last chapter of this section, contributed by Joshi and colleagues, discusses the uses of various molecular techniques and cutting-edge technologies to invent point-of-care diagnostics for various diseases in animals.

The last section of this book, "Bioinformatics, big data, and integrated omics," includes nine chapters that cover various perspectives on the biocomputational approach and study of integrated omics to unveil the system biology of livestock *vis-à-vis* artificial intelligence and big data analysis in managing and improving animal farms. Chapter 16 "Bioinformatics: unveiling the systems biology" stressed the role of in silico methods in studying the systems biology of animals. Various biocomputational tools used to decipher biological information have been discussed here. The 17th chapter, written by Debajit Dey, discusses "Bioinformatics in the development of antivirals," where various bioinformatics interventions have been discussed. This chapter is of immense importance regarding the application point of view in virology and diagnostics. Chapter 18 by Mukhopadhyay and Jadoun converse various areas of applications of artificial intelligence in animal health. The chapter is beneficial in the era of artificial intelligence and robotics. The next chapter, "Computational genomics approaches for livestock improvement and management,"

elaborates on the diverse application of genomics and computational biology to improve livestock health. Chapter 20 has been written by Manisha Malhotra and team deals with the basic concept of artificial intelligence and machine learning and their applications in livestock management and precision farming. Kashyap and Deshmukh have contributed to Chap. 21 where the role of the internet of things has been elaborated with possible detailed applications in animal sciences. It is the next step toward utilizing sensors, mechatronics, and robotics for innovative animal management. These two chapters are practically instrumental. Chapters 22, 23, and 24 cover the most pertinent and very advanced perspectives in today's time. E-Agriculture, as discussed in the second last chapter, holds excellent promise for animal sell-purchase in a more thoughtful and better way. The last chapter deals with quantum computing, a burning topic in today's scenario.

The contents of the book are a reader's digest to update and upgrade one's knowledge about recent perspectives of biotechnological and biocomputational advancements in various domains of animal sciences. The authors of these book chapters are highly qualified scientists in disease diagnosis, genomics, metagenomics, microbial genomics, proteomics, stem cell biology, bioinformatics, and data science. The editors are grateful to all the contributors for their enthusiasm and cooperation during the book compilation, review, and revision process. We also thank our reviewers for their constructive advice. We also extend our thanks to the Springer team for their cooperation, from acceptance of the proposal to the production of this book. This book is aimed at scientists, researchers, and other professionals and is an invaluable and timely review of new and innovative biotechnological interventions that have the potential to enhance animal health and livestock production.

Ludhiana, Punjab, India

Chandra Sekhar Mukhopadhyay
Ratan Kumar Choudhary
Harsh Panwar
Yashpal Singh Malik

Acknowledgments

Publishing a book sounds like a “dream come true.” The people who are associated with the entire process deserve special mention and applause. The authors acknowledge all of those who have contributed to the book in one or another form. At the outset, our sincere thanks to all the authors for their contribution to the scientific society by pouring their latest knowledge and skills into the form of book chapters. The publishers also deserve a special note of thanks and regards. Besides, our sincere thanks to the reviewer of the chapters and their critical comments to give a proper edge to the readers of the book. The authorities of Guru Angad Dev Veterinary and Animal Sciences University deserve special mention for providing a conducive environment for its faculty to excel in the global arena through research, academic, and extension activities.

If anyway, we are missing anyone who has contributed to this book, our sincere apologies, and heartfelt thanks for their contribution.

Contents

Editors and Contributors

About the Editors

Chandra Sekhar Mukhopadhyay is a Senior Scientist at the College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana. He has 15 years of working experience in research and teaching UG, PG, and Ph.D. students in Bioinformatics, Genomics, and Animal Genetics. Dr. Mukhopadhyay has been engaged in research works involving animal genomics and transcriptomics directed toward the reproduction and disease resistance of cattle, buffalo, and chicken. He has earned three extra-murally funded research projects to his credit. He has undergone 6 months advanced training program on Marker-Assisted Selection and Genomics at Iowa State University (ISU), Ames, USA, during 2011–12. He is a Fulbright fellow (FNAPE) and has worked at ISU, in the area of biostatistical modeling of imprinted loci in bovines. He has authored a book on bioinformatics published by John Wiley and Sons, USA. He is one of the national-level trainers on genome-wide association studies and genomic selection in training programs in India. Dr. Mukhopadhyay and his team have worked on bubaline micro-RNAs associated with diseases. He is currently working on parentage determination in bovines and canines.

Ratan Kumar Choudhary is an Assistant Professor at the College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Punjab. Dr. Choudhary completed his doctoral degree in Animal Science from the University of Maryland, USA. His doctoral research work was aimed at bovine mammary stem cell characterization and in vivo cellular manipulation for enhanced lactation persistency in bovines. He worked at the United States Department of Agriculture (USDA) facility for more than 6 years (2007–2012). He has two post-doctoral research experiences, one from the University of Kentucky (2012–2013) and the other from the University of Vermont (2018–2020), USA. The long-term goal of his research is to develop stem cell therapies for dry period management, and mastitis in dairy animals and to develop regenerative medicine for companion animals.

Harsh Panwar is an Assistant Professor in Dairy Microbiology, at the College of Dairy Science and Technology, Guru Angad Dev Veterinary and Animal Sciences University, India. His research interests include antimicrobial resistance and its mitigation in dairy and food. Dr. Panwar has published more than 50 articles in high-impact journals. He has been awarded Young Scientist Award 2019 by the Association of Microbiologists of India, the Indo-Australian Career Boosting Gold Fellowship 2018–19 by the Department of Biotechnology, Best Teacher 2019, and the Best Researcher 2018 Award by Guru Angad Dev Veterinary and Animal Sciences University, DST Inspire Fellowship by the Department of Science and Technology, and University Gold Medal by Kurukshetra University Kurukshetra. Dr. Panwar has co-edited two books on the Mitigation of Antimicrobial Resistance published by Springer Nature and is an editorial board member of several journals of repute.

Yashpal Singh Malik is currently the Dean of the College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, India. He is a recipient of the prestigious "ICAR National Fellow" award while working at the ICAR-Indian Veterinary Research Institute. His areas of expertise are rotaviral diseases, viral disease epidemiology, microbial biodiversity, host–virus interactions, and pathogen diagnostics. He has pursued advanced studies in molecular virology at the University of Minnesota, USA; the University of Ottawa, Ontario, Canada; and the Wuhan Institute of Virology, Wuhan, China. He is the recipient of several prestigious awards and honors, including the ICAR-Jawaharlal Nehru Award. He has authored 9 books, 60 book chapters, and over 200 research and review articles. Prof. Malik is a member of the International Committee on Taxonomy of Viruses (ICTV) on the *Birnaviridae* and *Picobirnaviridae* study group. He has been associated with societies of international repute like the World Society for Virology (USA). He has been serving as Secretary General of the Indian Virological Society. Being a member of "One Health group in the Federation of Asian Veterinary Association" (FAVA) for 2021–2025, he is the Indian flag bearer on the international forum. Prof. Malik is the Editor-in-Chief of the Journal of Immunology Immunopathology and Associate Editor of Frontiers in Microbiology and Frontiers in Veterinary Sciences. He has been awarded the prestigious Fellowship by the Academy of Microbiological Sciences, the National Academy of Dairy Sciences, and the National Academy of Biological Sciences.

Contributors

Sheikh Firdous Ahmad Division of Animal Genetics, ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

Santosh Anand Department of Dairy Microbiology, Phulo-Jhano Murmu College of Dairy Technology, Birsa Agricultural University, Hansdiha, Dumka, Jharkhand, India

Palak Arora National Institute of Animal Biotechnology (NIAB), Hyderabad, Telangana, India

Luit Barkalita Department of Animal Biotechnology, Faculty of Veterinary Sciences, Assam Agricultural University, Guwahati, Assam, India

Gourab Basak Centre for One Health, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

Ashish Bedi School of Life and Environmental Sciences, Deakin University, Geelong, Victoria, Australia

Subhomoi Borkotoky Department of Biotechnology, Invertis University, Bareilly, Uttar Pradesh, India

Shikha Chaudhary Department of Bioinformatics, College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

Supriya Chhotaray Animal Genetics and Breeding Division, ICAR—National Dairy Research Institute, Karnal, Haryana, India

Ritwika Das Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, PUSA, New Delhi, India

Sarmistha Debbarma Department of Animal Resources Development, Government of Tripura, Agartala, Tripura, India

Dipak Deka Department of Animal Biotechnology, Faculty of Veterinary Sciences, Assam Agricultural University, Guwahati, Assam, India

Bharti Deshmukh Department of Animal Genetics, and Breeding, College of Veterinary Sciences, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

Debajit Dey School of Medicine, University of Maryland, Baltimore, MD, USA

Sunny Dholpuria Department of Life Sciences, J. C. Bose University of Science and Technology, YMCA, Faridabad, Haryana, India

Swati Dubey National Institute of Animal Biotechnology (NIAB), Hyderabad, Telangana, India

Sonu Gandhi National Institute of Animal Biotechnology (NIAB), Hyderabad, Telangana, India

Arundhati Wandhare Ganesh Department of Dairy Microbiology, College of Dairy Science and Technology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

Ablesh Gautam Central Research Institute, Kasauli, Himanchal Pradesh, India

Himani Gautam College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

Sandeep Goel National Institute of Animal Biotechnology (NIAB), Hyderabad, Telangana, India

Mir Asif Iquebal Division of Agricultural Bioinformatics (Earlier Centre for Agricultural Bioinformatics), ICAR-Indian Agricultural Statistics Research Institute, PUSA, New Delhi, India

Abhilash Jadhao Department of Veterinary Pathology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

Yogendra Singh Jadoun Department of Veterinary & A.H. Extension Education, College of Veterinary Science, Guru Angad Dev Veterinary and Animal Sciences University (GADVASU), Ludhiana, Punjab, India

Pranay Jain Department of Biotechnology, University Institute of Engineering and Technology, Kurukshetra University, Kurukshetra, Haryana, India

Sarika Jaiswal Division of Agricultural Bioinformatics (Earlier Centre for Agricultural Bioinformatics), ICAR-Indian Agricultural Statistics Research Institute, PUSA, New Delhi, India

Akanksha Jaiswar Division of Agricultural Bioinformatics (Earlier Centre for Agricultural Bioinformatics), ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India

Shalini Jaswal Animal Biotechnology Center, ICAR-National Dairy Research Institute, Karnal, Haryana, India

Vinay G. Joshi Department of Animal Biotechnology, College of Veterinary Sciences, Lala Lajpat Rai University of Veterinary and Animal Sciences (LUVAS), Hisar, Haryana, India

Deepali Kalambhe Centre for One Health, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

Neeraj Kashyap Department of Bioinformatics, College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University (GADVASU), Ludhiana, Punjab, India

Bhawanpreet Kaur Department of Bioinformatics, College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

Geetika Kaur Post Doctorate Fellow, Integrative Biosciences Center, Wayne State University, Detroit, MI, USA
Department of Ophthalmology, Visual and Anatomical Sciences, School of Medicine, Wayne State University, Detroit, MI, USA

Navkiran Kaur Department of Veterinary & A.H. Extension Education, College of Veterinary Science, Guru Angad Dev Veterinary and Animal Sciences University (GADVASU), Ludhiana, Punjab, India

Richa Khirbat Department of Animal Biotechnology, College of Veterinary Sciences, Lala Lajpat Rai University of Veterinary and Animal Sciences (LUVAS), Hisar, Haryana, India

Amit Kumar Division of Animal Genetics, ICAR-Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, India

Dinesh Kumar Division of Agricultural Bioinformatics (Earlier Centre for Agricultural Bioinformatics), ICAR-Indian Agricultural Statistics Research Institute (IASRI), PUSA, New Delhi, India
Department of Biotechnology, School of Interdisciplinary and Applied Sciences, Central University of Haryana, Mahendergarh, Haryana, India

Sudarshan Kumar Animal Biotechnology Center, ICAR-National Dairy Research Institute, Karnal, Haryana, India

Wenli Li USDA-ARS, Cell Wall Biology and Utilization Research, Madison, WI, USA

K. M. Lokesh Centre for One Health, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

Sushila Maan Department of Animal Biotechnology, College of Veterinary Sciences, Lala Lajpat Rai University of Veterinary and Animal Sciences (LUVAS), Hisar, Haryana, India

Manisha Malhotra Department of Botany, Banaras Hindu University, Varanasi, Uttar Pradesh, India

Yashpal Singh Malik College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

Basavaraj Mathapati Indian Council of Medical Research—National Institute of Virology, Dibrugarh Unit, Dibrugarh, Assam, India

Prabhakar Maurya CEHTRA (Fr) (India Office), New Delhi, India

Dwijesh Chandra Mishra Centre for Agricultural Bioinformatics (CABin), ICAR-Indian Agricultural Statistics Research Institute (IASRI), Indian Council of Agricultural Research, New Delhi, India

Chandra Sekhar Mukhopadhyay College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

Ram Prakash Pandey Department of Biotechnology, University Institute of Biotechnology, Chandigarh University, Mohali, Punjab, India

Harsh Panwar College of Dairy Science and Technology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

Drishya Prakashan National Institute of Animal Biotechnology (NIAB), Hyderabad, Telangana, India

Ram Kumar Pundir Department of Biotechnology, University Institute of Biotechnology, Chandigarh University, Mohali, Punjab, India

Anil Rai Division of Agricultural Bioinformatics (Earlier Centre for Agricultural Bioinformatics), ICAR-Indian Agricultural Statistics Research Institute, PUSA, New Delhi, India

Nivedita Rai School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India

Shivam Saini National Institute of Animal Biotechnology (NIAB), Hyderabad, Telangana, India

Vikas Sangwan Department of Dairy Microbiology, College of Dairy Science and Technology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

Anindita Sarma Department of Animal Biotechnology, Faculty of Veterinary Sciences, Assam Agricultural University, Guwahati, Assam, India

Maitri Shah National Institute of Animal Biotechnology (NIAB), Hyderabad, Telangana, India

Ankita Sharma Central Research Institute, Kasauli, Himanchal Pradesh, India

Abhimati Shukla Department of Biochemical Engineering, Harcourt Butler Technical University, Kanpur, India

Swati Shukla National Institute of Animal Biotechnology (NIAB), Hyderabad, Telangana, India

Manvesh Kumar Sihag Department of Dairy Chemistry, College of Dairy Science and Technology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

Meena Sindhu Department of Microbiology, College of Basic Sciences and Humanities, Chaudhary Charan Singh Haryana Agricultural University, Hisar, Haryana, India

Amandeep Singh Directorate of Extension Education, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

Anamika Singh Department of Dairy Microbiology, College of Dairy Science and Technology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

Kalpana Singh Division of Agricultural Bioinformatics (Earlier Centre for Agricultural Bioinformatics), ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India

Kumar Siddharth Singh Institute for Microbiology, Leibniz University, Hannover, Germany

Mahavir Singh College Central Laboratory, College of Veterinary Sciences, LUVAS, Hisar, Haryana, India

Namita Ashish Singh Department of Microbiology, Mohanlal Sukhadia University, Udaipur, Rajasthan, India

Nikhlesh K. Singh Department of Ophthalmology, Visual and Anatomical Sciences, School of Medicine, Wayne State University, Detroit, MI, USA

Sumeet Singh Department of Veterinary Microbiology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

Vivek Kumar Singh Department of Veterinary Physiology and Biochemistry, College of Veterinary Sciences and Animal Husbandry, Junagarh, Gujarat, India

Jupi Talukdar Mahisah Perspicientia LLP, Guwahati, Assam, India

G. Taru Sharma National Institute of Animal Biotechnology (NIAB), Hyderabad, Telangana, India

Manish Tiwari Department of Animal Biotechnology, College of Veterinary Sciences, LUVAS, Hisar, Haryana, India

Geetika Verma School of Medicine and Health Sciences, University of North Dakota, USA

Vikas Vohra Animal Genetics and Breeding Division, ICAR—National Dairy Research Institute, Karnal, Haryana, India

Part I
Functional Genomics and Proteomics

Chapter 1
Genome Editing: Applications in Animal Health

Basavaraj Mathapati and Vivek Kumar Singh

Abstract Genome editing is the precise manipulation of genetic material in a living organism by deleting, replacing, or inserting a DNA sequence to alter the cell's fate or organism's traits. It typically aims to improve a crop or farm animal or to correct a genetic disorder. Genome editing involves creating a double-strand break (DSB) using specific nuclease systems, joining the DNA ends precisely through a homologous repair mechanism or a nonhomologous end joining (NHEJ). The advancement in biotechnological tools and newer genome editing platforms has paved the way for more precise and efficient alterations into genomes. Apart from introducing genetic modifications for desirable traits in animals, gene editing tools have also been implied in the development of diagnostic, prevention (vaccine development through genome editing in infectious pathogens), and therapeutic modalities for improving animal health. The present chapter elaborates on the applications of modern biotechnology tools of "genome editing" in preserving animal health through various means like selective breeding, development diagnostics, therapeutics, and other disease prevention modalities like effective vaccines.

Keywords Genome editing · Animal health · Animal husbandry · Livestock · Editing platforms · CRISPR-Cas

1.1 Introduction

Animal husbandry and livestock contribute 40% of the global value of agricultural product output and support the livelihood of around 1.3 billion global population in terms of food and nutrition security and resources (FAO). The World Bank, an

B. Mathapati (✉)
Indian Council of Medical Research—National Institute of Virology, Dibrugarh Unit, Dibrugarh, Assam, India

V. K. Singh
Department of Veterinary Physiology and Biochemistry, College of Veterinary Sciences and Animal Husbandry, Junagarh, Gujarat, India

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_1

international financial institution under United Nations Organization (UNO), considers the livestock sector a pillar of the global food system and a major contributor to agricultural development, food security, and poverty reduction. Livestock also acts as a valuable asset to the majority of vulnerable populations in medium and low-income countries where they are used as draft animals owing to less mechanization in agriculture.

Livestock waste, like manure, is an essential source of natural fertilizer. Hence, in some geographical regions, livestock production systems have been viewed as potential contributors to the preservation of biodiversity and to carbon sequestration in soils and biomass. In difficult environments such as drylands and mountains, livestock is considered a key factor that sustainably converts natural resources into food and fiber and work power for local communities. Apart from this, these animals are also part of a larger sphere of human utilities ranging from companion partners as pets, research entities as laboratory animals, and defender squads in security agencies. Wildlife as a whole is a key element of the forest ecosystem, adding to the recreational activities of humans. In animals, health may be defined as normal physiological functioning and normal behavior as per the defined norms and criteria for each species, in simple terms, the absence of disease. In terms of production, animal health is a status that allows the highest productivity. In more general terms, animal health is a balance between the animal, its environment, and its physical welfare. Anything which can affect this balance results in disease; moreover, animal health focuses on prevention rather than cure. Preservation of animal health status involves two components; first, making the animals less vulnerable to any causal factor that affects health which includes making them genetically fit in prevailing conditions and following the management practices that boost health. Second, making them resilient to causal factors by acquired means which includes vaccines against infectious microorganisms and, to a lesser extent, early therapeutic cures when health is derailed. The advent of modern biotechnology has proved to be efficient in addressing both components of preserving animal health.

1.2 Genome Editing

1.2.1 Historic Perspective

Alterations in the genomes were studied with the help of spontaneous mutations in the 1920s. In later years, it was demonstrated that the rate of mutagenesis could be enhanced by radiation and the use of some chemicals (Muller 1927; Auerbach et al. 1947). Down the timeline, transposable elements were relied on for insertional mutagenesis and transgenesis; the cut-and-paste DNA transposons that translocate by excision from a donor site and insertion into a target site were used as tools in genetic engineering. All these procedures, though, could produce stable alterations in genomes, but were at random sites in the genome. Later in the late 1970s and early 1980s first, every attempt was made for targeted genome editing in eukaryotes like

yeasts and later in mice (Scherer and Davis 1979; Rothstein 1983; Smithies et al. 1985; Thomas et al. 1986). All these targeted genome editing technologies relied on the homologous recombination (HR) process and were precise, but the efficiency was very low in mammalian cells and also included the tedious task of selection and meticulous characterization to yield desired results. Owing to these disadvantages along with low frequency, these methods were not suitable for other mammalian species. The present-day genome editing platforms have overcome some of these difficulties and are used in genetic manipulations in all kinds of cells and organisms.

1.2.2 Present-Day Genome Editing Technologies

The present-day high-efficiency genome editing technologies rely on the ability to make targeted DNA-double-strand breaks (DSBs) in the DNA target of interest. The editing technologies were developed by a thorough understanding of mechanisms of DNA damage and repair, which take place in cells naturally. Process such as meiosis involves recombination between homologous sequences; it was also known that DSBs could be generated by ionizing radiations. Researchers discovered highly specific nucleases that are known to stimulate the events in homologous repair in yeast and mammalian cells. These developments paved the way forward for programmable genome editing. Later developments in the more precise homology-independent joining of broken ends through a process called nonhomologous end joining (NHEJ) brought pace to be new and efficient genome editing technologies. At present, three genome editing platforms are very promising and have been widely used. These platforms rely on the ability of a class of nucleases which can be programmed to create DSBs at any desired site in a DNA. These are zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats-associated system (CRISPR-Cas). The genome editing modalities are shown in the following figure (Fig. 1.1).

1.2.2.1 Zinc-Finger Nuclease-Based Editing Platforms

ZFNs are named because the zinc-binding repeats were used as DNA-binding domains in the sequence-specific transcription factor that was first to be characterized for the action of these classes of nucleases. The ZFNs consist of two main domains, a cleavage domain from bacterial protein and a set of zinc-finger motifs which were identified in sequence-specific transcription factors from eukaryotes. The peptide modules are made to interact with base pair triplets. The DNA-recognition specificity can be altered by altering the few residues in a single zinc finger, and these alterations can be devised to recognize many different DNA triplets.

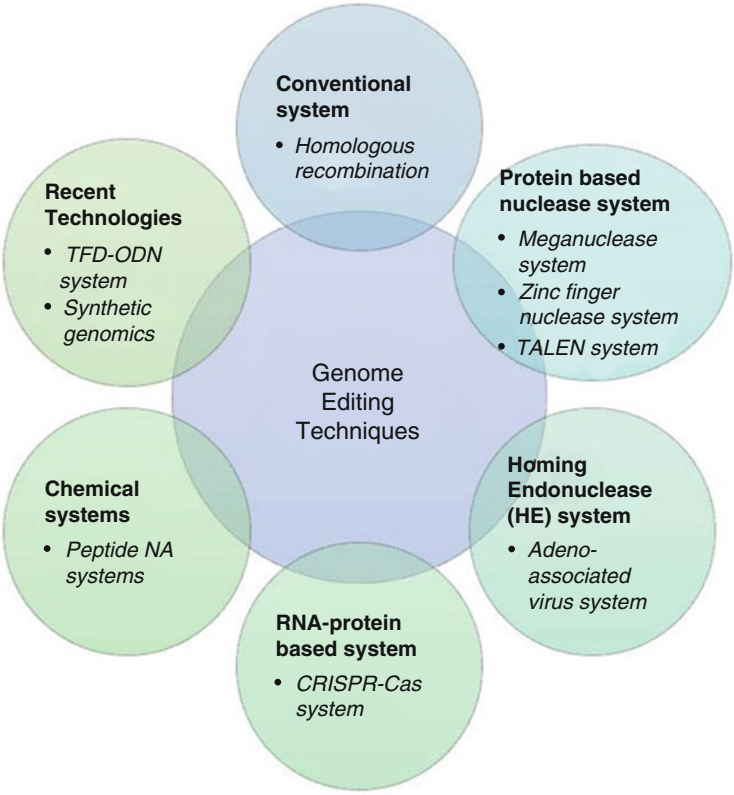

Fig. 1.1 Genome editing technologies

1.2.2.2 TALENs-Based Editing Platforms

TALENs consists of cleavage domains from bacteria, as in ZFNs and DNA recognition modules from transcription factors of plant-pathogenic bacteria. Many plant-pathogenic bacteria secrete proteins into host cells capable of regulating host genes to promote infection by binding to specific sequences. A good number of bacterial restriction enzymes cut the DNA a few bases away from the sites they recognize, as they have separate binding and cleavage domains. The bacterial cleavage domains do not have inherent sequence specificity. Hence, these can be linked to specific and novel DNA-binding domains for precise DNA editing. Many zinc finger arrays and TALE arrays have been designed to generate ZFNs and TALENs.

1.2.2.3 CRISPR-Cas

The most widely used genome editing tool by researchers is CRISPR-associated proteins (Cas), which are encoded adjacent to the clustered regularly interspaced

short palindromic repeats (CRISPR); initially, they were shown to match some viral genomes and were identified as a defense modality in bacteria against viruses. These repeat clusters capture the invading viral genomes, and the process is guided by short RNAs called crRNAs which are encoded in the CRISPR arrays. The crRNAs are processed by a system of small trans-acting RNA (tracrRNA) and participate in the cleavage activity of invading DNA which was characterized first in Streptococcus pyogenes. The mix of crRNA, tracrRNA, and a cleavage system (Cas9) can be used for editing genomes with better efficiency and ease.

1.3 Applications of Genome Editing in Animal Health

Genome editing platforms based on ZFNs and TALENs are mostly used in the research arena of agriculture and medicine, and CRISPR-Cas-based platforms are very popular in recent years and preferred by research laboratories across the globe for genome editing. Genome editing is a powerful tool to assess individual gene functions and precisely manipulate the DNA to produce the desired alteration in cell behavior or functions. By virtue of the ability to carry out specific modifications in the genes responsible for a specific function or regulatory role, one can bring in or introduce new genes (knock-in) or delete the undesirable ones (knockout) for the direct benefit of the host. Apart from it, genome editing can be a handy tool in understanding the pathogenesis of many diseases like cancer, genetic disorders, and a few infectious diseases as well. Applications of genome editing can be under three broad areas (Fig. 1.2).

The most beneficial applications of genome editing were through a process called gene drive in which introduced genetic material/element in an organism is rapidly spread through a breeding population by copying its genetic element into the organisms which they previously did not have. Such gene drives were part of the biosphere and existed in some populations, but genome editing technologies like CRISPR-Cas9 have provided a way to create synthetic or artificial gene drives. Many CRISPR-Cas9-based artificial gene drives have successfully been introduced into mosquito populations which are vectors for many diseases of humans and animals. Genome editing has created a gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae* (Hammond et al. 2016). Similarly, a gene drive has been created that efficiently inactivates the genes crucially essential for parasite growth in mosquitoes through CRISPR-Cas9-based genome editing (Gantz et al. 2015). The introduction of such a population of mosquitoes in the natural ecosystem has shown to be effective in reducing the disease burden. The approach of creating gene drives through genome editing in vector populations can have a tremendous impact on the control or elimination of a few vector-borne diseases in animals and humans (Carroll 2017). Releasing organisms of modified genetics has evoked many appropriate concerns about unpredictable effects on nature as a whole; even similar concerns are with genetically modified organisms (GMOs). These apprehensions are slowly being reduced as

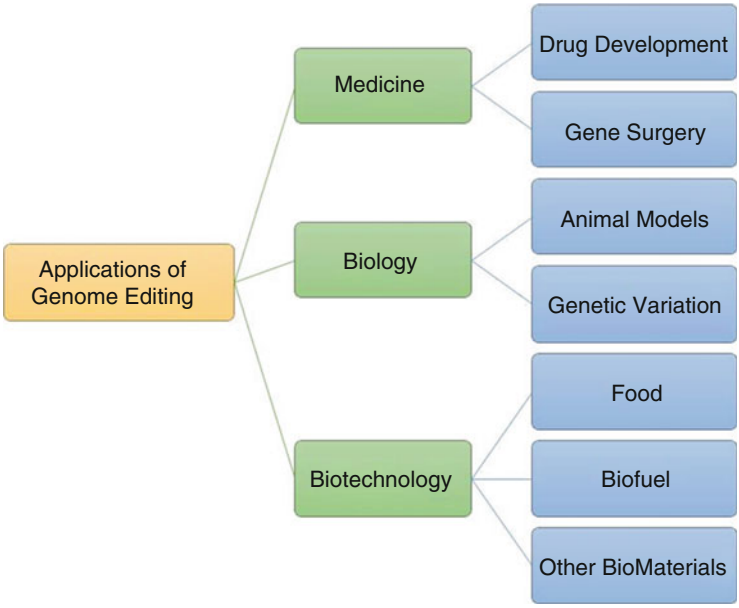

Fig. 1.2 General applications of genome editing

the genome editing platforms can be used for purposes that do not involve the introduction of genetic material from one species into another, but modifying the naturally existing genetic material at a very precise location and current day high-throughput sequencing platforms can provide whole genome sequences to look into off-target mutations. In agriculture, disease-resistant varieties of wheat (Zhang et al. 2017), potatoes with better keeping quality (Clasen et al. 2016), and soy plants with modified oil composition in seeds (Demorest et al. 2016) are a few examples of applications of genome editing.

The specific applications in animal sciences (Fig. 1.3) include the development of hornless cattle, which can avoid painful dehorning and prevent injuries while transporting and in-fighting (Carlson et al. 2016). The hornless cattle were bred selectively by gene manipulation through TALENs. The improved platforms for genome editing in sheep and cattle have demonstrated efficient modes for the production of animals with desirable economic traits, including disease resistance (Proudfoot et al. 2015). Similarly, the researchers have developed swine resistant to Porcine Reproductive and Respiratory Syndrome (a viral disease caused by PRRSV of the *Arteriviridae* family) by suppressing the production of a protein called CD163, which is a crucial component of infection within the pigs that the virus uses for spread, using a CRISPR-cas-based genome editing platform (Whitworth et al. 2016; Burkard et al. 2018). The pigs with double knockout (through CRISPR-cas) of both genes for pAPN and CD163 were shown to resist the PRRSV and transmissible gastroenteritis virus (TGEV) infections (Xu et al. 2020). Another deadly disease of pigs caused by the African swine fever virus is being tackled

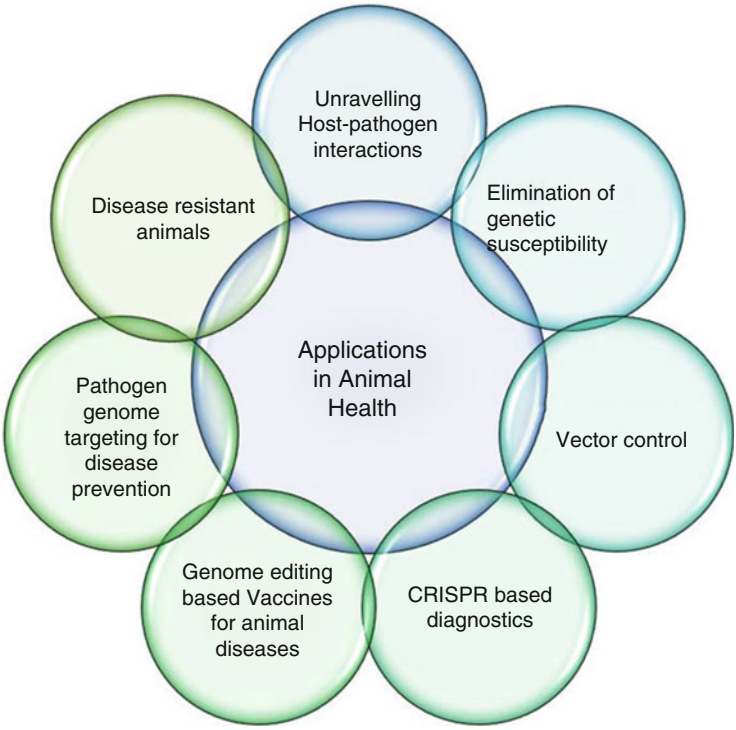

Fig. 1.3 Applications of genome editing in animal health

through genome editing efforts by modifying a specific gene in domestic pigs by its counterpart in warthogs in which disease is not manifested and not fatal, through the ZFNs platform of gene editing. Genome-edited domestic pigs with counterpart genes from warthogs have been shown to lessen the severity and viral load of ASFV (McCleary et al. 2020), but did not prevent the infection. African swine fever resistance in pigs was demonstrated experimentally through a CRISPR-Cas9-mediated genome editing platform targeting viral genes (Hubner et al. 2018).

Research efforts are also taken up to produce and breed the poultry by introducing a "decoy" molecule that can interrupt the replication and spread of avian influenza viruses, thereby restricting the transmission. Researchers have used the TALENs-based approach to introduce a few genes of mice into cattle to improve resistance against tuberculosis (Tuggle and Waters 2015). The modified cattle possesses immune cells that are efficient in slowing down disease progression and making them less susceptible to developing overt symptoms of the disease.

The genome editing platforms have also provided newer diagnostics and vaccines for the control of infectious diseases in animals and humans. In recent times nucleic acid-based diagnosis of infectious diseases has gained popularity. The PCR-based nucleic acid detection techniques are predominant diagnostic modalities that are used in disease surveillance and control programs. These PCR-based diagnostics are

cost-intensive and need specialized equipment through the advent of genome editing platforms like CRISPR, and nucleic acid-based diagnostic tools shaping up into point-of-care diagnostic modalities where the need for specialized equipment is minimized. Several CRISPR-based diagnostic tools are employed in the surveillance and monitoring of livestock diseases. Modalities like SHERLOCK (specific high-sensitivity enzymatic reporter unlocking), HOLMES (<1 hr, low-cost multipurpose highly efficient system), and DETECTR (DNA endonuclease-targeted CRISPR trans reporter) are popular examples of CRISPR-based viral detection methods. These modalities rely on the trans-cleavage potential of the Cas12a and Cas13 aproteins. Point-of-care or on-site diagnostic tools for pseudorabies, African swine fever, PRRS, and Japanese encephalitis based on CRISPR technology have been developed recently, adding to animal health globally (Sollner et al. 2021). CRISPR-based editing technologies have been used for creating desired mutations in various viral vectors, including the herpes virus of turkey, infectious laryngotracheitis virus, and duck enteritis virus. These modified vectors are edited to carry the genes and express specific antigens as viral vector vaccines against many diseases. New Castle disease, Marek's disease, infectious laryngotracheitis, pseudorabies, and avian influenza are a few examples of avian diseases where the vaccines based on genome editing have shown promising results (Vilela et al. 2020).

1.4 Conclusion

Genome editing is one of the most powerful tools in modern-day biotechnology, which has a wide range of applications. It enables the introduction/deletion/replacement of genetic elements in the genomes precisely to modify the phenotypic traits as desired. It has revolutionized many areas of modern biomedical science. The genome alteration that began with targeted mutagenesis in the early 1970s has eventually reached the production of disease-resistant animals and plants using genome editing platforms developed over the years. Genome editing technologies are being used in drug development, understanding the host-pathogen interactions, the establishment of animal models to address infectious and noninfectious diseases in humans, production of plants and animals with desirable production traits, development of low-cost diagnostics and vaccines for human and animal diseases, generation of biomaterials, and treatment of genetic disorders and cancer. In animal health, genome editing has played a vital role in the alteration of genetic susceptibility to specific infections, reduction of vector-borne disease burden through vector elimination or alterations at population levels, low-cost pen-side diagnostic tools for animal disease surveillance and monitoring, and affordable vaccines for prevention and control of important animal diseases. However, genome editing technologies are well accepted in all spheres of applications, barring the alterations of genomes of animals and vectors per se, which evoked appropriate concerns linked to undesired genetic alterations in hosts and unpredictable potential long-term impacts on ecosystems. The robust present-day high-throughput sequencing technologies and

bioinformatic tools have enabled us to check over the undesired genetic alterations throughout the genomes after manipulation. As genome editing technologies progress, many different applications can be perceived, but responsible adjustments can be brought in the present-day practices so as to make these advancements in genome editing technologies safe, effective, and favorable to mankind across the globe.

References

Auerbach C, Robson JM, Carr JG (1947) Chemical production of mutations. Science 105:243–247

Burkard C, Opriessnig T, Mileham AJ, Stadejek T, Ait-Ali T, Lillico SG, Whitelaw CBA, Archibald AL (2018) Pigs lacking the scavenger receptor cysteine-rich domain 5 of CD163 are resistant to porcine reproductive and respiratory syndrome virus 1 infection. J Virol 92(16): e00415–e00418

Carlson DF, Lancto CA, Zang B, Kim ES, Walton M, Oldeschulte D, Seabury C, Sonstegard TS, Fahrenkrug SC (2016) Production of hornless dairy cattle from genome-edited cell lines. Nat Biotechnol 34(5):479–481. https://doi.org/10.1038/nbt.3560

Carroll D (2017) Genome editing: past, present, and future. Yale J Biol Med 90(4):653–659

Clasen BM, Stoddard TJ, Luo S, Demorest ZL, Li J, Cedrone F et al (2016) Improving cold storage and processing traits in potato through targeted gene knockout. Plant Biotechnol J 14(1): 169–176

Demorest ZL, Coffman A, Baltes NJ, Stoddard TJ, ClasenBM LS 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):225

Gantz VM, Jasinskiene N, Tatarenkova O, Fazekas A, Macias VM, Bier E et al (2015) Highly efficient Cas9-mediated gene drive for population modification of the malaria vectormosquito *Anopheles stephensi*. Proc Natl Acad Sci U S A 112(49):E6736–E6743

Hammond A, Galizi R, Kyrou K, Simoni A, Siniscalchi C, Katsanos D et al (2016) A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. Nat Biotechnol 34(1):78–83

Hubner A, Petersen B, Keil GM et al (2018) Efficient inhibition of African swine fever virus replication by CRISPR/Cas9 targeting of the viral p30 gene (CP204L). Sci Rep 8:1449. https://doi.org/10.1038/s41598-018-19626-1

McCleary S, Strong R, McCarthy RR, Edwards JC, Howes EL, Stevens LM, Sánchez-Cordón PJ, Núñez A, Watson S, Mileham AJ, Lillico SG, Tait-Burkard C, Proudfoot C, Ballantyne M, Whitelaw C, Steinbach F, Crooke HR (2020) Substitution of warthog NF-κB motifs into RELA of domestic pigs is not sufficient to confer resilience to African swine fever virus. Sci Rep 10(1): 8951. https://doi.org/10.1038/s41598-020-65808-1

Muller HJ (1927) Artificial transmutation of the gene. Science 66:84–87

Proudfoot C, Carlson DF, Huddart R, Long CR, Pryor JH, King TJ et al (2015) Genome edited sheep and cattle. Transgenic Res 24(1):147–153

Rothstein RJ (1983) One-step gene disruption in yeast. Methods Enzymol 101:202–211

Scherer S, Davis RW (1979) Replacement of chromosome segments with altered DNA sequences constructed in vitro. Proc Natl Acad Sci U S A 76:4951–4955

Smithies O, Gregg RG, Boggs SS, Koralewski MA, Kucherlapati RS (1985) Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature 317(6034):230–234

Sollner J-H, Mettenleiter TC, Petersen B (2021) Genome editing strategies to protect livestock from viral infections. Viruses 13:1996. https://doi.org/10.3390/v13101996

Thomas KR, Folger KR, Capecchi MR (1986) High frequency targeting of genes to specific sites in the mammalian genome. Cell 44:419–428

Tuggle CK, Waters WR (2015) Tuberculosis-resistant transgenic cattle. Proc Natl Acad Sci U S A 112(13):3854–3855. https://doi.org/10.1073/pnas.1502972112

Vilela J, Rohaim MA, Munir M (2020) Application of CRISPR/Cas9 in understanding avian viruses and developing poultry vaccines. Front Cell Infect Microbiol 10:581504. https://doi.org/10.3389/fcimb.2020.581504

Whitworth KM, Rowland RR, Ewen CL, Trible BR, Kerrigan MA, Cino-Ozuna AG, Samuel MS, Lightner JE, McLaren DG, Mileham AJ, Wells KD, Prather RS (2016) Gene-edited pigs are protected from porcine reproductive and respiratory syndrome virus. Nat Biotechnol 34:20–22. https://doi.org/10.1038/nbt.3434. PMID: 26641533

Xu K, Zhou Y, Mu Y, Liu Z, Hou S, Xiong Y, Fang L, Ge C, Wei Y, Zhang X, Xu C, Che J, Fan Z, Xiang G, Guo J, Shang H, Li H, Xiao S, Li J, Li K (2020) *CD163* and *pAPN* double-knockout pigs are resistant to PRRSV and TGEV and exhibit decreased susceptibility to PDCoV while maintaining normal production performance. *ELife* 9:e57132. https://doi.org/10.7554/eLife.57132

Zhang Y, Bai Y, Wu G, Zou S, Chen Y, Gao C et al (2017) Simultaneous modification of three homoeologs of TaEDR1 by genome editing enhances powdery mildew resistance in wheat. Plant J 91(4):714–724

Chapter 2
Stem Cell Therapy: Promises and Challenges in Treating Animal Diseases

Sandeep Goel, Sonu Gandhi, Swati Dubey, Maitri Shah, Shivam Saini, Palak Arora, Drishya Prakashan, Swati Shukla, and G. Taru Sharma

Abstract Regenerative medicine and emerging biotechnologies stand to revolutionize the practice of modern medicine. Stem cell therapy holds great promise in treating many incurable diseases in veterinary medicine. Stem cells are undifferentiated cells that can self-renew and differentiate into several types of cells with specific functions. Caplan pioneered the concept of stem cell therapy in 1991. Since then, it has gained considerable attention since the use of stem cells overcomes several shortcomings associated with traditional therapies. The genesis of stem cells is followed by laboratory steps of controlled stem cell culturing and derivation. There is a wide range of stem cells, including embryonic stem (ES) cells, induced pluripotent stem cells (iPSCs), and adult stem cells. Even though ES cells and iPSCs are almost identical practically in many biological attributes, iPSCs are favored over ES cells because iPSCs are devoid of religious and ethical concerns, and autogenous iPSCs are immune-free, unlike ES cells. Tissue regeneration has become a clinical reality due to advancements in stem cell biology, including embryonic and postnatal somatic stem cells. Stem cells are widely used in therapeutic applications in veterinary medicine to treat several diseases, including musculoskeletal injuries and wounds. Because of their simple isolation and in vitro culture techniques with no ethical issues, mesenchymal stem cells (MSCs) appear most suitable for therapeutic applications. MSCs are gaining traction in veterinary medicine due to their unusual immunomodulatory activities. However, stem cell therapy still faces numerous challenges, including a low long-term cell survival rate after transplantation, a short homeostasis maintenance period of blood glucose, immunological rejection, and tumorigenesis. Recent developments in stem cell research are primarily driven by the limitations of current treatment options for various medical problems in animals. Because of their versatility, the employment of scaffolds and extracellular vesicle-based therapies needs special consideration among the many types of stem tissue applications. Biodegradable scaffolds are

S. Goel · S. Gandhi · S. Dubey · M. Shah · S. Saini · P. Arora · D. Prakashan · S. Shukla · G. Taru Sharma (✉)
National Institute of Animal Biotechnology (NIAB), Hyderabad, Telangana, India
e-mail: tarusharma@niab.org.in

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_2

essential for generating a 3D environment that promotes tissue development. The use of scaffolding materials combined with stem cell technology has excellent tissue regeneration potential. Many biomaterial scaffolds comprising synthetic, natural, and ceramic materials have been employed in regenerative medicine. This study focuses on the most commonly used stem cell types, emphasizing MSCs and their clinical applications in cell-based therapies in veterinary medicine and the challenges that stem cell therapy must overcome to be accepted worldwide.

Keywords Regenerative medicine · Stem cells · Veterinary medicine · 3D stem cell culture · Scaffolds · Synthetic biomaterials

2.1 Introduction

Since their discovery, stem cells have been regarded as the cornerstone of regenerative medicine and tissue engineering in veterinary medicine. Stem cells are undifferentiated cells that can self-renew and differentiate into various cell types. Their source classifies them as embryonic stem (ES) cells and adult stem cells and their potency as totipotent, pluripotent, multipotent, and unipotent stem cells (Barky et al. 2017), in which adult stem cells include mesenchymal and spermatogonial stem cells (SSCs) as shown in Fig. 2.1. In addition, induced pluripotent stem cells (iPSCs) are also considered stem cells that are generated by reprogramming adult or

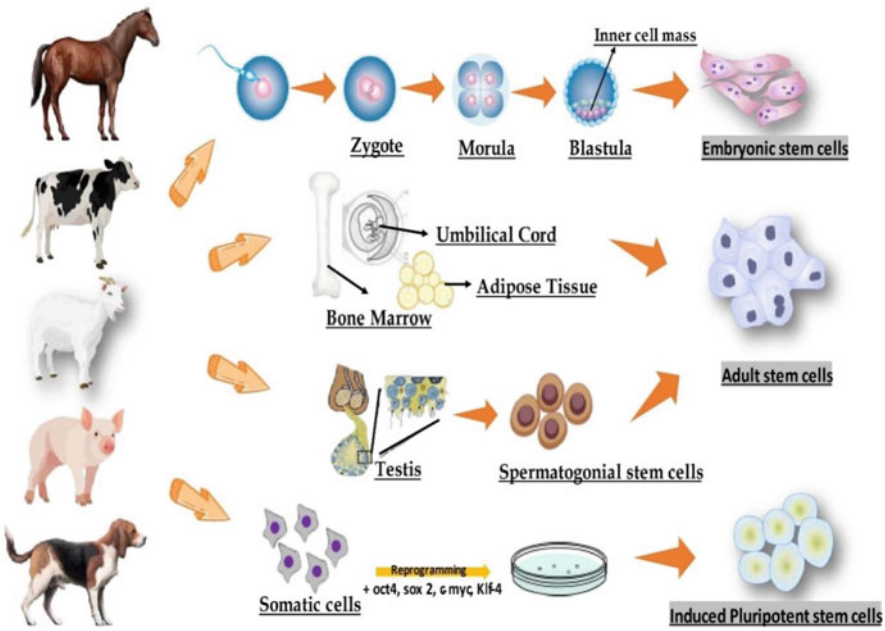

Fig. 2.1 Different types of stem cells used in veterinary medicine for therapy

fetal somatic cells by expressing pluripotent genes. Martin Evans (Nobel Prize 2007) and Matt Kauffman were the first to identify, isolate, and successfully culture ES cells using mouse blastocysts in 1981. ES cells were used because they offered a high-potential cell source for genome editing to create genetically modified organisms. However, as highly efficient gene-editing tools such as CRISPR/Cas9 and somatic cell nuclear transfer (SCNT) became available, ES cells applications shrank dramatically. In addition, iPSCs were created by reprogramming mouse embryonic fibroblasts to provide an alternative to ES cells (Kumar et al. 2021).iPSCs have emerged as a promising alternative to overcome the ethical and immunogenic challenges associated with ES cells. However, due to their tumorigenicity, iPSCs were unsuccessful in stem cell therapies. Mesenchymal stem cells (MSCs), on the other hand, have a significant advantage over ES cells and iPSCs in regenerative medicine, making them a more appealing candidate for stem cell-based therapy (Hwang et al. 2009). Adult stem cells, such as MSCs, have recently seen a considerable increase in use, owing to their strong regeneration potential and no ethical issues, to treat various pathological and clinical disorders that are otherwise difficult to cure in animals. MSCs are multipotent undifferentiated stem cells that may be harvested and cultured in vitro. They can be found at various sites throughout an animal's body. In veterinary sciences, most stem cell therapy has been established using bone marrow and adipose-derived MSCs and has been employed in livestock, domestic, and companion animals to treat various illnesses and clinical conditions. For instance, wound healing (Martinello et al. 2018), canine spinal cord injuries (Bhat et al. 2019), and horse tendon injuries (Smith et al. 2003), to name a few.

Moreover, MSC-secreted trophic factors and MSC-impregnated scaffolds are frequently used in animals for wound healing and musculoskeletal defects (Sun et al. 2013). Even though MSCs are a viable candidate for stem cell therapy because of their immunomodulatory, anti-inflammatory, and antimicrobial capabilities, the high expansion cost and local storage facilities remain essential obstacles to the clinical implication of stem cell therapy in veterinary medicine. The lack of standard criteria to define MSCs, an unclear image of their mechanism of action, and less preclinical data on the safety and efficacy of MSCs-based therapies are significant challenges to the clinical use of MSCs-based therapies in veterinary sciences (Devireddy et al. 2017). Moreover, many challenges, such as poor long-term maintenance of stem cell function, low cell retention, and engraftment, limit the successful use of stem cell translation into clinical practice. Constructive regeneration can be performed using extracellular matrix (ECM) scaffolds and stem cells because the cells can be delivered to the site of the infraction and adhere once delivered, instead of being "lost." Because of the niche-like conditions of ECM, stem cells tend to differentiate into tissue-specific cells and attain several characteristics like functional cells even in the absence of any directed differentiation using external inducers. Therefore, a supportive microenvironment needs to regulate stem cell function by stimulating underlying host repairment after cell administration.

2.2 Need for Stem Cell Therapies

There are no or inefficient therapies for many diseases, including neurological problems, spinal cord injuries, diabetes, etc. Even if a remedy exists, it comes with negative consequences. For example, antibiotic therapies in bovine mastitis cause fibrosis, leading to decreased milk production and poor animal health. For that, alternative remedies are necessary to overcome the limits of traditional therapy. Furthermore, treating diseased tissue early with MSCs might ameliorate or even cure the disease, minimizing the need for life-long treatment, improving the patient's quality of life, and perhaps lowering overall societal expenditures (Oreffo et al. 2005). Most notably, MSCs aid in reducing microbial load and inflammation at the site of infection. They also repair the damaged tissue, reducing the danger of reinfection and restoring the patient's health. With the growing demand for animal-based products throughout the world, poor animal health has a detrimental influence on both production and the quality of animal-based products, which negatively impacts the economy in the livestock sector.

Despite available therapies increasing zoonotic diseases and antimicrobial resistance, diseases persist in animals for prolonged periods. Many stem cell-based therapies, including MSCs-based therapies, have been developed for diseases such as wound healing, bone and cartilage defects, and musculoskeletal disorders in animals to overcome the limitations of existing treatment and provide a superior alternative (Chandra et al. 2022). However, most of them are either in the early stages of preclinical or clinical research. Even though significant exploratory work has been done in animals to develop stem cell-based therapy, numerous effectiveness issues persist. Immunomodulation and in vivo models of action remain unanswered. There are few stem cell-based products available in the market for the treatment. Thus, there is a significant gap between the demand and availability of economic and practical stem cell-based therapies or products.

2.3 Type of Stem Cells Used for Cell Therapy

2.3.1 Embryonic Stem (ES) Cells

ES cells are defined as pluripotent stem cells derived from the blastocyst's inner cell mass, which have the potential to self-renew and can differentiate into cells of three germ layers, therefore capable of forming all types of cells and tissues in an adult body. ES cells are known to express many pluripotent markers, including Oct4, Sox2, and Nanog. The first isolation and characterization of ES cells were reported from mouse embryos in 1981 (Martin 1981), followed by rhesus monkey embryos in 1995 (Thomson et al. 1995) and then human embryos in 1998 (Thomson et al. 1998). These findings triggered extensive research in the veterinary world to establish ES cell lines from various domestic animals and elucidate their mechanism of

action in regenerative medicine and tissue engineering. Researchers have made several attempts to establish ES cell lines from domestic or companion animals in the past few decades, but still, there is no stable ES cell line to date. However, some ES cells have been isolated and cultured in vivo from pigs and bovine; as these cells did not follow the standards of real ES cell lines, they are being called ES-like cell lines (Hou et al. 2016; Bogliotti et al. 2018). Earlier, ES cells were considered a powerful tool for gene editing in tissue engineering. However, with the development of highly efficient gene-editing tools such as CRISPR/Cas9 and SCNT, the ES cells-based therapeutic practices gradually decreased in veterinary medicine (Wilmut et al. 1997).

Moreover, despite immense differentiation and proliferation potential with high regenerative ability, ES stem cell-based therapies remained restricted to reproductive applications and were not successful in regenerative medicine as they cause tumorigenesis at the site of transplantation. These tumorigenic properties of ES cells make them unsuitable for grafting or transplantation at the site of infection or wound. In addition, another major limitation associated with ES cells is ethical issues to establish the cell lines.

2.3.2 Induced Pluripotent Stem Cells (iPSCs)

The therapeutic approach in regenerative medicine generally requires many cells, which restricts ES cells in domestic animals because of the lack of standard isolation protocol and availability of stable ES-cell lines. Here, iPSCs show significant therapeutic potential with self-renewal capacity and vast differentiation potential. These pluripotent stem cells are generated by reprogramming somatic cells by ectopic expression of pluripotent-specific transcription factors, with a thought to provide alternative stem cell therapy to overcome the limitations associated with ES cells-based therapies in regenerative medicine. First iPSCs generation was reported in 2006 when scientists had identified the four primary transcription factors responsible for maintaining pluripotency and transdifferentiating potential of ES cells, i.e., Oct3/4, Sox2, c-Myc, and Klf-4, had been ectopically expressed in mouse embryonic and adult fibroblast cells, followed by iPSCs derivation from human dermal fibroblast. These endogenous pluripotent factors transformed the terminally differentiated somatic cells into ES-like cells called iPSCs. Since their discovery, the iPSCs have been successfully generated from domestic and companion animals to develop large animal models to study human diseases and treat various chronic pathological or wound conditions that are otherwise difficult to cure using existing conventional therapies in animals. In domestic animals, the first isolation and derivation of iPSCs were reported from porcine, followed by a canine, equine, bovine, and many other species. After derivation of iPSCs from porcine, they have been differentiated into the cells of various lineages such as neural progenitor cells (Kim et al. 2019a), endothelial cells (Wei et al. 2020), and vascular smooth muscle

cells (VSMC) (Luo et al. 2017). Later, VSMC was augmented onto the scaffold to generate 3D scaffold-free tissue rings.

Similarly, equine iPSCs also have been differentiated into several cells and tissues, including neurons (Sharma et al. 2014), tendons (Bavin et al. 2015), and osteoblasts (Baird et al. 2018). Musculoskeletal injuries are considered the most common defects in horses; therefore, artificial tendons have been attempted to derive from iPSCs. However, only the 2D assay showed the matrix contraction and appropriate gene expression, while the 3D assay failed (Bavin et al. 2015). The same work extended to canines, where iPSCs were differentiated into chondrocytes and osteocytes using hydrogel culture conditions, which can be further used to treat musculoskeletal disorders (Whitworth et al. 2014). Initially, bovine iPSCs were generated using fetal or skin fibroblast by viral transduction methods. These cell lines were not stable even after transduction and could not be grown indefinitely (Han et al. 2011). Some bovine iPSCs were generated successfully with stable characters; however, the stability of the cell line seemed to be controlled by exogenous transcription factors.

Moreover, the significant applications of iPSCs were restricted to the generation of germline transgenic/chimaera animals and to providing various types of differentiated cells for drug modelling and screening as an animal model. Although iPSCs hold great regenerative potential, they become less significant in regenerative medicine due to the risk of high tumorigenicity and immunogenicity. iPSCs are known to form teratomas at the transplantation site and generate a high immune response, so there are fewer iPSCs-based clinical applications in domestic and companion animals.

2.3.3 Adult Stem Cells

2.3.3.1 Spermatogonial Stem Cells (SSCs)

In adult male vertebrates, spermatogenesis is a continuous germ cell proliferation, differentiation, and maturation that occurs throughout their lives. These stem cells divide asymmetrically, producing both stem cell population and progenitor cells, eventually giving rise to adult spermatozoa via spermatogenesis. SSCs have gotten much interest in recent decades because of their ability to develop transgenic animals and preserve the germplasm of domestic, companion, and wild animals. The preservation of germplasm is essential in conserving endangered wild species, where SSCs can be isolated, maintained, and transplanted to host testis to form the haploid gametes (Goel et al. 2011). The continuous research on SSC biology resulted in various approaches to biomanipulation, which have led to the development of novel biotechnological solutions in reproductive biology and regenerative medicine. Enough SSCs are required to investigate their potential, biomanipulation, and stem cell therapies. Hence, isolation, characterization, and in vitro expansion of SSCs are done based on SSC-specific markers to acquire a suitable number of cells. Genetic

manipulations of isolated germline stem cells and subsequent transplantation to produce transgenic spermatozoa are two of the most common applications of SSC-based therapy in domestic and companion animals. In animals where ES cells are unavailable, and SCNT has various issues, transgenesis using male germline cells is a superior approach.

Furthermore, SSCs transplantation within or between species provides an excellent way to preserve the reproductive potential of genetically valuable domestic and wild animal species. In addition, infertility has long been regarded as a severe issue for farm animals, resulting in financial loss and a decline in high-yield breeding pools. By transplanting the germ cells extracted from a fertile donor animal and commencing donor-derived spermatogenesis in the recipient, SSCs transplantation has emerged as a promising alternative to ES cells for restoring the recipient's reproductive ability. However, SSCs are the only adult stem cells that are capable of transferring the genetic material from one generation to the next, also, at the same time, having the potential to convert into pluripotent stem cells (Guan et al. 2006). The initial work of SSCs transplantation was done in mice and further extended to farm and wild animals. To date, SSCs have been used for many applications in farm animals to target various aspects, including the production of transgenic farm animals by transplanting genetically altered male germ cells (Niemann and Kues 2007), development of transgenic pigs to provide tissues and organs for xenotransplantation to humans (Niemann and Kues 2003), and production of therapeutic proteins in the milk of dairy animals (Keefer 2004).

2.3.3.2 Mesenchymal Stem Cells (MSCs)

MSCs were discovered initially by Friedenstein in the 1960s as bone-forming cells in the bone marrow of guinea pigs (Friedenstein et al. 1966), and Owen expanded such work to rats (Luria et al. 1987). Since their discovery, stem cell therapy has gained significant attention and has been used in preclinical and clinical research for many years (Fig. 2.2). MSCs are defined as undifferentiated adult multipotent cells that can self-renew and differentiate into cells of other lineages (Pittenger et al. 1999). In the past few decades, researchers have explored and discovered several properties of MSCs, and perhaps some are yet to be discovered. Unlike hematopoietic stem (HSCs) cells originating from bone marrow, MSCs can be isolated from other sources, including placenta, adipose tissue, teeth, menstrual fluid, dental pulp, and umbilical cord (Hass et al. 2011), as shown in Fig. 2.2. MSCs are easy to isolate, grow readily in the culture dish, have intrinsic differentiation potentials, produce an abundance of helpful growth factors cytokines, be genetically modified, and are immune-evasive, which permits use in allogenic conditions (Caplan 2015). The International Society for Cellular Therapy (ISCT) defined MSCs as functionally based upon negative (CD45, CD34, CD14/CD11b, CD19/CD20/CD79α, and HLA-DR) and positive (CD73, CD90, and CD105) cell surface markers, plastic adherence, and trilineage differentiation (Canham et al. 2010; Dominici et al. 2006). However, these criteria were defined for human MSCs; hence its expression profile

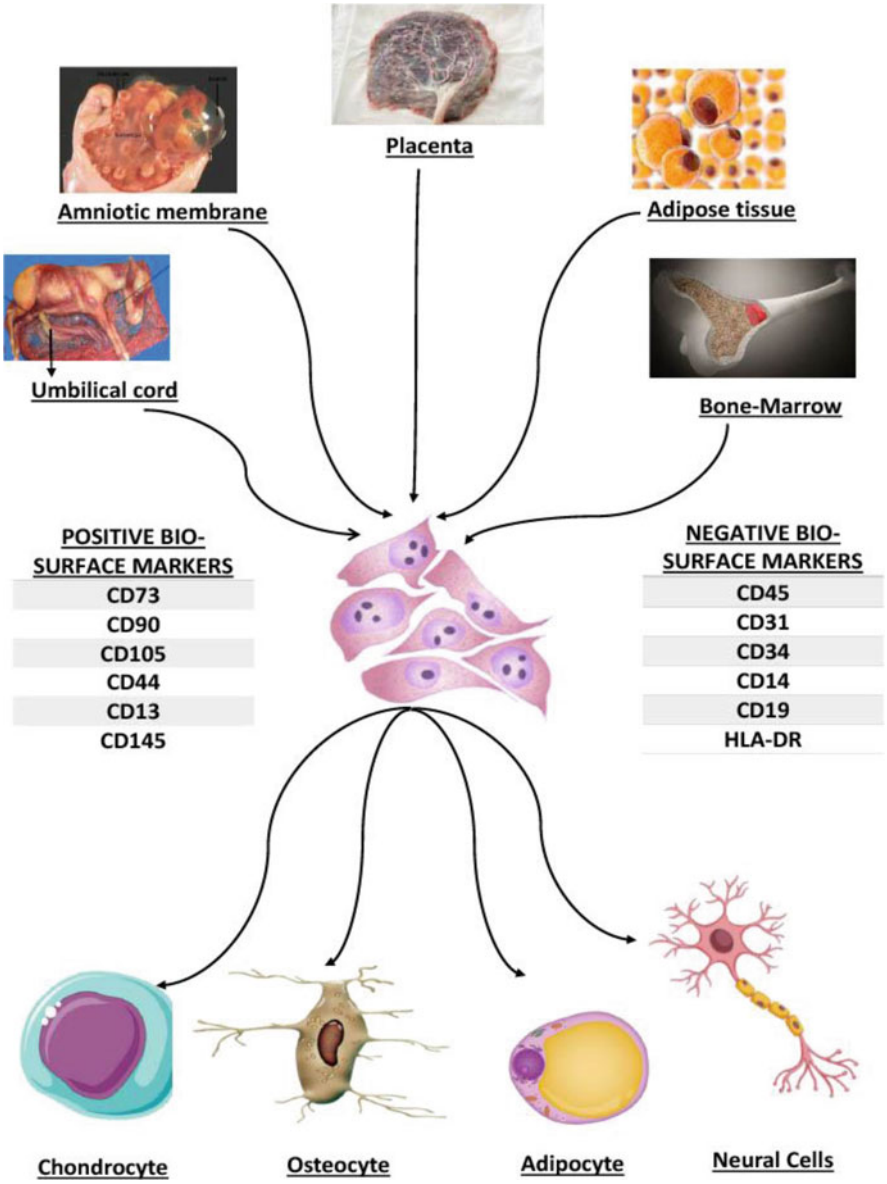

Fig. 2.2 Origin, characterization, and differentiation of mesenchymal stem cells

may vary in different animal models, posing confusion among the researchers. Studies suggest that activated MSCs secrete exosomes in response to stimulatory signals due to various pathological conditions, which further can be used for cell therapies using MSCs conditioned media. Exosomes are extracellular vesicles (ECV) discovered in sheep reticulocytes in 1983 (Harding et al. 1983); due to the

phospholipid bilayer on their surface, exosomes have good stability and permeability and can be secreted by almost all cell types (Villarroya-Beltri et al. 2014). The magnificent therapeutic potential of MSCs, primarily associated with their capabilities of initiating tissue regeneration and inhibiting inflammation, makes them suitable for stem cell therapy. Owing to their high regenerative potential, MSCs are widely used in tissue engineering and regenerative medicine in veterinary sciences. Several diseases in livestock remain difficult to treat due to the limitations associated with existing treatments and management of the disease; therefore, MSCs-based therapies offer a great alternative to conventional therapies in livestock and companion animals. This shows the great promise of MSCs in regenerative medicine; however, isolation, in vitro culture, and storage remain an expensive task to make them available for their clinical applications.

2.4 Current Status of MSC Research and its Outcome

There has been growing interest in applying MSC transplantation in regenerative medicine to repair and restore the normal function of diseased and injured tissues. MSCs have also been shown to have immunosuppressive and healing capacities, improve angiogenesis, and prevent fibrosis (Liu et al. 2010). Several animal studies have shown that MSC treatment improves wound healing in goats (Pratheesh et al. 2017), sheep (Gomiero et al. 2018), horses (Textor et al. 2018), and dogs (Johnson et al. 2017). The use MSCs has been proposed as a 'magic bullet' approach toward skeletal tissue regeneration (Djouad et al. 2009). Bone marrow MSCs (BM-MSCs) have been used in treating diseases in animal models and human clinical trials in areas such as autoimmune diseases, solid organ allograft survival, hepatic cirrhosis, kidney diseases, fracture healing, neuro- and muscle-degenerative diseases, myocardial infarction, and spinal cord injury (Ikehara 2010; Popp et al. 2009; Dai et al. 2009; Chhabra and Brayman 2009; Kitada and Dezawa 2009; Peer et al. 2022). Neurological recovery has been shown in animal models of Parkinson's disease, hypoxic-ischemic neural damage, and retinal injury following in vivo transplantation of these cells inside the lesion (Bouchez et al. 2008). Spinal cord injuries are among the most common injuries in humans and animals. The results of studies using MSCs to treat traumatic spinal cord injuries showed some positive effects, but more research is needed to find a way to supplement the therapeutic effects of MSC therapies (Bhat et al. 2019). MSCs have shown the potential to improve the healing of skin defects in animal models (Satoh et al. 2004; Wu et al. 2007; Chen et al. 2008) and humans. MSC-derived ECV injected locally to treat circular wounds in dogs resulted in significantly improved cutaneous wound healing (El-Tookhy et al. 2017). MSCs can home into sites of injury, where they modulate the repair process. The development of bioengineered skin grafts seems to hold promising potential in tissue regeneration in human and animal clinical practice, given certain obvious advantages, especially due to rapid and almost scar-free healing. MSC-based treatment has been largely successful in treating the symptoms of autoimmune diabetes in animal

models. Administration of MSCs therapy to cows with mastitis caused by *S. aureus* resulted in a lower bacterial count in milk (Peralta et al. 2020). Stem cell therapy is also being researched in ophthalmology. Subconjunctival administration of autologous BM-MSC improved immune-mediated keratitis in three out of four horses, as evidenced by decreased neovascularization of the area, decreased surface irregularities, and increased corneal clarity (Davis et al. 2019). Similarly, Sgrignoli et al. found that 6 months after repeated topical administration of allogeneic adipose tissue-derived MSCs (AD-MSCs) into the conjunctival sac, CD4, IL-6, IL-1, and tumor necrosis factor were significantly reduced in dogs (Sgrignoli et al. 2019). Also, MSCs are being used in recurrent airway obstruction in horses (Barussi et al. 2016), while in cats, there are attempts to use them in digestive system disease (inflammatory bowel disease) and chronic kidney disease (Quimby et al. 2013; Quimby et al. 2011). Experimentally, in mice, rats, and dogs, the possibilities of using MSCs to treat acute liver failure were evaluated (Bhat et al. 2019; Banas et al. 2008). Moreover, auto- and allogeneic stem cells have been used in the experimental treatment of spinal cord injuries (Jung et al. 2009; Ryu et al. 2012), induced urinary incontinence (Strasser et al. 2006), mucosal ulcerations (El-Menoufy et al. 2010), muscular dystrophies (Nitahara-Kasahara et al. 2012), bone defects (Bigham-Sadegh et al. 2012), articular cartilage (Min et al. 2007), and diabetes (Zhu et al. 2011). Musculoskeletal disorders now account for most clinical cases in veterinary medicine where stem cells have been used. Osteoarthritis, tendon and ligament injury (primarily in horses), and intervertebral disc disease are the most common musculoskeletal diseases (Prządka et al. 2021). So far, surgical techniques, systemic and local application of anti-inflammatory preparations, hyaluronic acid, specialized cellular products (platelet-rich plasma, interleukin-1 receptor antagonist protein), movement restriction, and physiotherapy have been used to treat these diseases (Kirkby Shaw et al. 2020; Magri et al. 2019). Conventional treatment of musculoskeletal injuries involving articular cartilage, ligaments, and menisci damage is frequently associated with a poor prognosis for horse athletic performance. It was demonstrated by Smith et al. that autologous BM-MSC treatment of naturally occurring tendinopathies induces the formation of tissue resembling a normal tendon matrix rather than a fibrous tissue that is formed during the natural healing process (Smith et al. 2013). In addition to the autologous MSC therapy, promising results were reported with allogeneic MSC therapy for tendon and ligament disorders such as tendinitis of superficial and deep digital flexor tendons and desmitis of the suspensory and inferior check ligaments (Van Loon et al. 2014). Moreover, oral pain and mastication problems can have a major impact on the quality of the animal's life. In addition to usual sources of stem cells such as bone marrow and adipose tissue, cells derived from local tissues such as dental pulp stem cells (Abdelaz et al. 2019; Çolpak et al. 2019) or periodontal ligament stem cells are studied as a therapeutic option in oro-dental diseases (Shi et al. 2018). In the mini-pig periodontal defect model, allogeneic AD-MSCs alone can induce periodontal tissue regeneration (Venkataiah et al. 2019).

2.5 Types of MSCs Used for Therapies

2.5.1 Bone Marrow-Derived Mesenchymal Stem Cells (BM-MSCs)

Bone marrow is the primary source of adult stem cells in research, where cells are isolated by bone marrow aspiration and further can be used for cell-based therapy (Fig. 2.3). A BM-MSCs have been leading stem cell research in regenerative medicine (Aithal et al. 2021). It is one of the critical class MSCs for the following reasons (1) they facilitate the activity of HSCs, (2) BMSCs could be employed to promote extramedullary tissue healing, and (3) they support bone development and remodeling. They give rise to osteoblasts, which deposit osteocalcin in mineralized bone tissue, and they can control osteoclasts, which facilitate the release of osteocalcin, an osteogenic marker whose expression is regulated by the RUNX2 gene (Shabir et al. 2022). These BM-MSCs are found in higher volumes within the trabeculae of the marrow cavities and, for isolation, require prior consent from the ethical committee. BM-MSCs can produce growth factors and bioactive molecules which act as paracrine signaling molecules (Heldring et al. 2015). Several MSC-based therapies are currently either in the preclinical or clinical trial stage to treat bone and cartilage damage (Jayaram et al. 2019; Dubey et al. 2018), spinal cord injury (Muniswami et al. 2018), neurodegenerative disease (Mahendru et al. 2021), cutaneous wound healing (Ansari et al. 2013; Bharti et al. 2020), liver and kidney

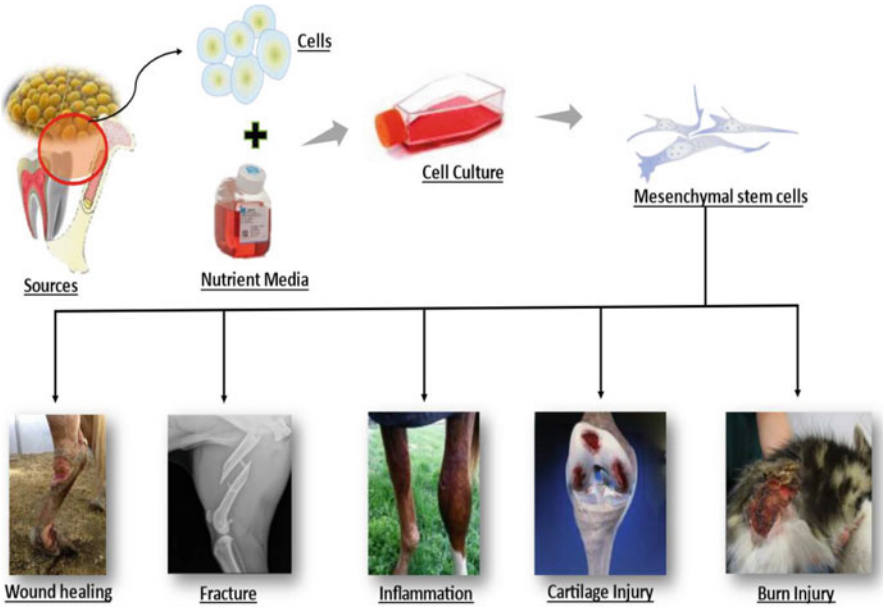

Fig. 2.3 Mesenchymal stem cells utilized for different therapies in veterinary medicine

disease, and diabetic mellitus-associated diseases (Bist et al. 2021). BM-MSC-derived exosomes are also being analyzed for disease-specific applications (Raghav et al. 2021).

2.5.2 Adipose-Derived Mesenchymal Stem Cells (AD-MSCs)

Over the past few years, it has been observed that fat is the energy reservoir and a good source of multipotent stem cells known as AD-MSCs. Isolation of MSCs from adipose tissues is much easier and less invasive (by liposuction) than from bone marrow. Also, these cells hold no ethical concern and can be isolated in the appropriate amount for in vitro culture and cell therapy.AD-MSCs are immunomodulatory and anti-inflammatory, which exclude the chances of any immune reaction while autologous or allogeneic transplantation in the host (Dabrowska et al. 2021). Inflammation is the first reaction at the site of infection or injuries, which causes tissue damage or sometimes severe conditions such as mastitis and wound injuries in diabetic patients. Recently, MSCs have been used to treat various pathological or metabolic disorders, such as wound healing (Mankuzhy et al. 2021), spinal cord injuries (Sharun et al. 2020), lateral sclerosis (Prabhakar et al. 2012), bone defects (Jeyaraman et al. 2021), liver pathology (Baligar et al. 2017), and diabetes (Vanikar et al. 2010). All studies have found a significant recovery and tissue damage repair compared to the existing conventional treatment.

2.5.3 Fetal Cord and Amniotic Membrane-Derived Mesenchymal Stem Cells

Stem cells isolated from fetus-associated amniotic membrane, umbilical cord, and umbilical cord blood have captured attention in the last decade in regenerative medicine. Umbilical cord blood-derived stem cells are generally HSCs that have proven to cure many hematological disorders in recent years (Udalamaththa et al. 2020). The umbilical cord and amniotic membrane are the ethical issue-free tissue collected for scientific research. The umbilical cord establishes a link between the developing fetus and the placenta. Earlier, the umbilical cord after parturition was considered of no medical use, but in recent times, it has drawn attention towards MSCs-based therapy in regenerative medicine. However, umbilical cord-derived mesenchymal stem cells (UC-MSCs) therapeutic potential has not been much explored in veterinary medicine earlier. Recently, UC-MSCs have been used for intra-articular stem cell therapy to treat chronic elbow arthritis and have shown a significant effect in alleviating lameness caused by osteoarthritis in dogs (Kim et al. 2019b). In addition to the umbilical cord, the amniotic membrane is also considered a good source of stem cells and can be used as a biomaterial for various clinical

applications. Treatment of various kinds of wounds and skin diseases is among the most common clinical applications of amniotic membranes and stem cells derived from the amniotic membrane in veterinary medicine (Murphy et al. 2020; Mankuzhy et al. 2021).

2.6 Stem Cell-Based Therapies for Various Diseases in Animals

In veterinary sciences, many diseases exist which have no or least effective treatment available in conventional medicine. Several conditions, including bone and cartilage injuries, skin diseases and wounds, mastitis, and spinal cord injuries, take significant time to cure using existing treatments, leading to animals' declining health status and productivity. Therefore, stem cells offer potential therapeutics that can cure diseases faster than currently available treatments and help tissue regeneration, bringing them back to a normal functional state. MSCs have been reported to efficiently treat various diseases, including wound healing, spinal cord injuries, musculoskeletal defects, bone defects, and cartilage injuries in animals, as shown in Fig. 2.3.

2.6.1 Musculoskeletal Disorders

In sports and companion animals, tendon and ligament injuries, joint injuries, and spinal cord injuries are frequently observed as musculoskeletal diseases that impair animals' locomotor and neuromuscular abilities. These traumatic injuries can naturally heal with scar formation, which affects the normal function of healed tissue. However, it compromises locomotor function and makes the animal more vulnerable to reinjuries. However, traditional tendon and ligament injury therapies involve cooling, bandaging, and rehabilitation with controlled exercise, requiring surgical interventions (Petrov et al. 2003). Nonetheless, these approaches cannot provide complete tissue healing, reinjury is common, and animals are often unable to return to their pre-injury level.

Moreover, stem cell-based therapies-mediated healing is faster and aids in regenerating damaged tissue and restoring the normal functionality of animals. To date, stem cell-based therapies have been employed to cure several such diseases, and in most cases, significant clinical improvement has been observed. In horses, tendon and ligament injuries are the most common musculoskeletal defect that adversely affects locomotor function and overall health (Dakin et al. 2011). Sometimes, the condition worsens and animals die.

The first report of autologous BM-MSC implantation into the horse digital flexor tendon was published in 2003 (Smith et al. 2003). BM-MSC treatment of naturally occurring tendinopathy helps restore the tissue to resemble a normal tendon matrix,

unlike normal tendon fibrous tissue (Smith et al. 2013). In addition to injuries, osteoarthritis is also considered another prevented illness in horses. Due to hypocellularity and avascularization, cartilage tissues have minimal ability to self-repair. In addition, the enormous loading force and mechanical stress on the articular surface during the performance aggravate the situation and lead to the ending of the athletic career of sports animals (Frisbie and Stewart 2011). Osteoarthrosis has been treated using autologous BM-MSCs as well as allogenic AD-MSCs. In both cases, a substantial improvement and reduction in lameness was observed after the administration of MSCs. Similarly, in dogs, spinal cord injuries are among the most common disorders which lead to lifetime impairment of locomotive and neuromuscular ability of the animals. Moreover, allogeneic BM-MSCs injected at the lesion site have shown remarkable recovery in vertebral compression, and after a few months, animals recover entirely (Sharun et al. 2020).

2.6.2 Wound Healing

Any disruption to the natural structure and integrity of the skin results in the formation of a wound, which must be healed within a short time; if failed, wounds become chronic and difficult to treat. In domestic and companion animals, several wounds are caused by skin burns, bites, physical injuries, and pathological infection, which take a significant time to heal naturally. However, scar formation and chances of reinfections are typical. Wound healing is a complex process that includes the interplay of many different growth factors, cytokines, and paracrine factors. Any misregulation of these secretory factors may lead to the formation of a chronic wound which takes a long time to recover, and proper management is required since open wounds (acute or chronic) are usually difficult to treat in animals due to inadequate wound management and vigilance, limitations of traditional approaches, and routine activity of animals.

In most cases, wounds gradually deteriorate the animal's health and harm productivity. Due to various constraints of traditional methods, stem cell therapies provide a great alternative to treat various wounds in less time. Moreover, adult stem cells such as MSCs hold an outstanding regenerative potential and healing ability as these cells help in cell recruitment by secreting bioactive factors, differentiation, immunomodulation, antimicrobial peptide secretion, angiogenesis, and reepithelialization at the transplanted site (Huang et al. 2020). Furthermore, MSCs have been explored to evaluate their therapeutic potential in various animal models to treat skin wounds (Ochiai et al. 2017), thermal burn wounds (Bliley et al. 2016), diabetic wounds (Shrestha et al. 2013), and radiation wounds (Liu et al. 2018). In all the studies, MSCs accelerated the healing of acute and chronic wounds and restored the typical epidermal/dermal architecture with reduced scar formation. MSCs are thus a promising candidate in regenerative medicine and tissue remodeling and can be used to treat various types of wound conditions in large animals.

2.7 When Stem Cell-Based Therapy Met Scaffolds: The Synergistic Effect

Stem cell biologists are now collaborating with engineers to integrate the emerging tissue engineering technology to create an artificial microenvironment (commonly referred to as a scaffold) to boost the success of stem cell-based therapy. The scaffold mimics the natural niche (commonly, ECM) of a particular tissue type at the topological, mechanical, and biochemical levels, which facilitates the transplanted stem cells' survival, proliferation, differentiation, and organization, consequently helping the cells to create the desired tissue type. Since each tissue type has its specific ECM, an ideal scaffold should mimic the ECM of the target tissue in its native state, at least partially. Hence, the critical role of scaffolds in stem cell-based therapy is analogous to the functions of ECM in native tissues and is linked to their architectural, mechanical, and biochemical characteristics (Fig. 2.4) (Chan and Leong 2008; Dubey et al. 2022).

For instance, architectural features, including topology, geometry, and porosity, have affected stem cell behavior. A variety of scaffold architectural features, including fiber length and diameter, aligned/interwoven patterns, and surface roughness (e.g., ridges, grooves, holes, and pillars), have been studied to provide stem cells with architectural signals to modulate their behavior towards different lineages (Viswanathan et al. 2015; Papadimitriou et al. 2020; Vijayavenkataraman 2020). The studies revealed that different topographies influenced cell shape, signaling pathways, and subsequent differentiation.

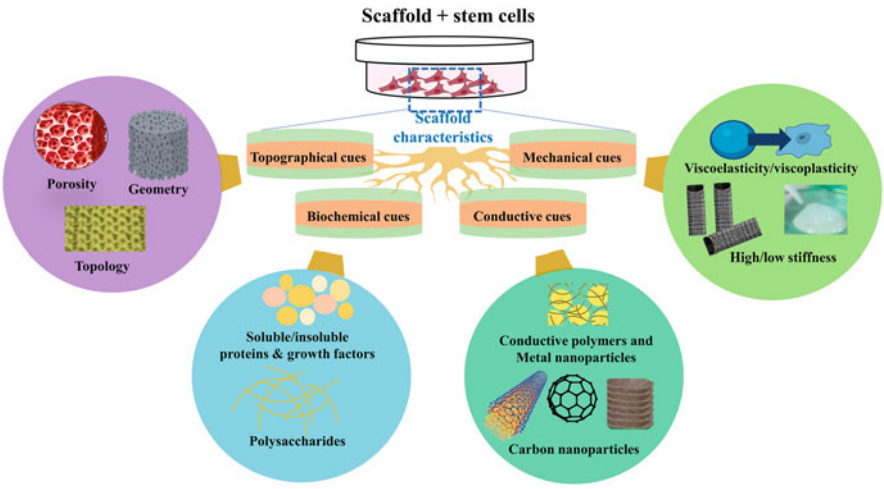

Fig. 2.4 Consideration of various scaffold characteristics for the scaffold-based culture of stem cells

Further, since ECM stiffness plays a crucial role in regulating development, homeostasis, and regenerative processes (Chaudhuri et al. 2020), there is direct evidence and substantial correlative data that viscoelasticity and material stiffness are important design parameters for scaffolds used in regenerative medicine. Various studies have shown mechano-dependent differentiation of stem cells. For example, Saha and colleagues (Saha et al. 2008) created a hydrogel culture system with variable moduli (10–10^4 Pa) to investigate the effect of mechanical signals on adult NSCs. They noticed that hydrogels with stiffness comparable to brain tissue (100–500 Pa) favored neuronal while stiffer gels (1–10 kPa) favored glial differentiation.

From a biochemical standpoint, the ECM contains a variety of soluble and insoluble proteins, growth factors, and polysaccharides that provide the cellular microenvironment with a biochemical context. ECM can act in overt and covert signaling by binding cell surface receptors or presenting noncanonical growth factors (Hynes 2009; Gattazzo et al. 2014). ECM components have also been integrated into the scaffolds to provide an instructive microenvironment with specific biochemical cues to facilitate tissue regeneration (Baiguera et al. 2014).

Next, the function of electroactive cells and tissues (such as neurons and cardiomyocytes) has benefited from electrical stimulation. Electrical stimuli have been shown to promote neural differentiation, neurite length, polarization, and migration of neurons (Zhu et al. 2019). Hence, there is a focus on the use of conductive polymers and scaffolds for neuronal and cardiac differentiation. A variety of electroactive materials, including conductive polymers (e.g., poly (3,4-ethylene dioxythiophene), polyaniline, polypyrrole, polythiophene), metal nanoparticles, and carbon materials (e.g., graphene, carbon nanotubes), are being used for the purpose (Shi et al. 2007; Seil and Webster 2010; Scott et al. 2021).

2.7.1 Scaffold-Based Culture

Scaffolds are made up of a porous network that is cultivated with cells in a way that encourages nutrient delivery and waste exchange. Numerous natural and synthetic biomaterials have been evaluated for scaffolding in the form of hydrogels, micro and nanofibers, and micro and nanospheres. An ideal scaffold should be biocompatible, biodegradable, nontoxic, highly porous, have good mechanical properties, and have a cell-binding affinity. Generally, two kinds of biomaterials, natural and synthetic, have been used to fabricate scaffolds based on the origin of the material. Both types have their advantages and disadvantages over each other. Synthetic scaffolds can be fabricated with the desired architecture and provide high mechanical support to the cells; however, they lack biocompatibility and biodegradability. On the contrary, natural biomaterials provide excellent cell adhesion and growth.

2.7.1.1 Natural Polymers

Components of the ECM such as collagen, hyaluronic acids, fibrinogen, elastin, glycosaminoglycans, and ECM-like materials such as chitosan cellulose are used as natural polymers (Filippi et al. 2020; Dubey et al. 2021). Scaffolds made out of these polymers are bioactive, biocompatible, biodegradable, not cytotoxic, assist cells to bind to the receptor, and provide a microenvironment for cellular responses (Ochiai et al. 2017). ECM proteins have a unique influence on cell behavior and response; hence, using them or their combinations as biomaterials results in the generation of functional scaffolds (Hinderer et al. 2016). These natural polymers have been widely used to fabricate stem cells treatment.

2.7.1.2 Synthetic Polymers

Synthetic polymers are artificial materials widely used as scaffolds in 3D culture because of their high versatility, availability, mechanical properties, non-immunogenicity, and reproducibility. In addition, synthetic polymers can be easily processed and are flexible compared to natural polymers. Extensive studies are being carried out with biodegradable synthetic polymers, including polyglycolide and polylactides (Taib et al. 2022).

2.7.2 Fabrication Methods for 3D Scaffolds

Porous polymeric matrices have been used as surfaces for cell support, attachment, growth, and further differentiation and proliferation on or inside their structures using various scaffold manufacturing techniques, as shown in Fig. 2.5. Scaffolds with a random structure, variable pore sizes, and minimal pore connectivity can be created using a variety of procedures typically employed in tissue engineering. The most extensively used techniques are solvent casting, freeze-drying, phase inversion, fiber bonding, melt-based technologies, and high-pressure-based processes. Electrospinning has lately been actively researched for producing submicrometric fiber meshes for various tissue engineering purposes. In addition to the approaches above, controlled microfabrication procedures have been developed to create scaffolds with complicated geometries and predetermined architecture. Moreover, laser-assisted bioprinting (LAB) is a microfabrication method used in stem cell therapy and artificial tissue engineering (Tortorella et al. 2022).

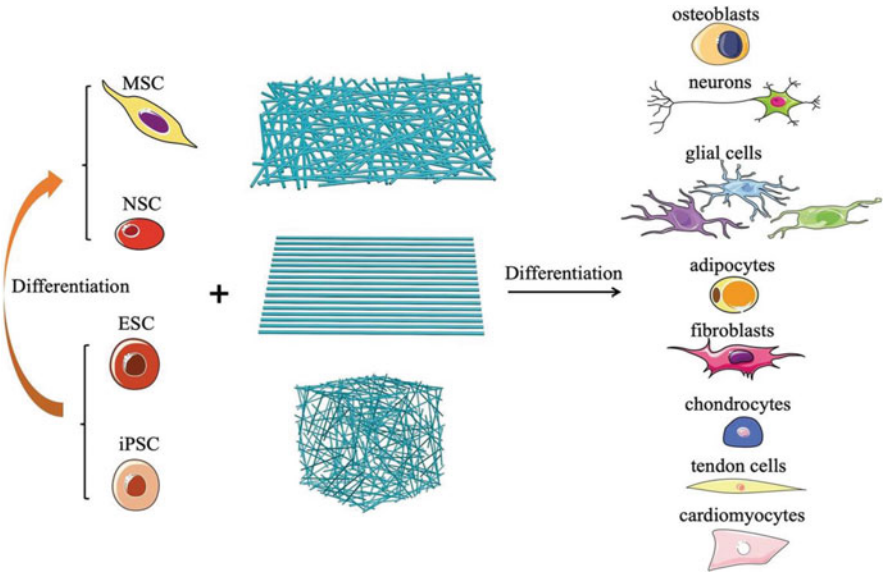

Fig. 2.5 Scaffolds serve as a platform to study the growth and differentiation of stem cells

2.7.3 Scaffold-Free Culture System

Scaffold-free 3D culture depends on the self-aggregation of cells in suspension culture (Langhans 2018). 3D spheroids recapitulate normal tissues' biological characteristics like cell-cell and cell-matrix interactions (Mueller-Klieser 1987). Spheroid-based stem cell treatment is a straightforward and effective way to boost stem cell therapeutic efficiency. As a result, this technique might be adapted to treat ischemia in patients. It was reported that when MSCs were cultured as 3D spheroids, the rate of cells surviving in ischemic tissues increased (Bhang et al. 2012). The microenvironment established within the spheroids favors the secretion of healing-inducing paracrine factors by UC-MSCs, thereby increasing elastin production and skin cell mobility (Santos et al. 2015). Another study in AD-MSCs spheroid showed that these spheroids led to tissue regeneration when transplanted into the wound bed in athymic mice. These findings suggest that spheroid transplantation is effective for stem cell therapy to treat wound beds (Park et al. 2015).

Furthermore, it was confirmed that 3D human placenta-derived MSCs secreted anti-inflammatory and trophic factors and enhanced angiogenesis and neurite morphogenesis in vitro *(*Deng et al. 2021*)*. The transplantation of human cord blood mononuclear cells as spheroids into mouse ischemic hindlimbs significantly improved the survival rate of the transplanted cells and significantly increased the number of micro vessels and smooth muscle α-actin-positive vessels (Bhang et al. 2012). AD-MSCs spheroids transported by a nanosheet have persisted on the skin wound surface for a long time, indicating that they might protect the transplanted

spheroids from the outside environment. Furthermore, AD-MSCs spheroids conveyed by the nanosheet stimulated wound healing and increased vascular endothelial growth factor(VEGF) and hepatocyte growth factor secretion, implying paracrine actions (Nagano et al. 2021). Neurosphere combined with fibrous scaffolds showed prominent differentiation into glial cells and neurons (Xue et al. 2020). Veterinary disorders can be treated using stem cells in the same way as humans can. Many clinics are now using autologous or allogeneic stem cell injections, fresh or grown in the lab, to treat various veterinary disorders. Adult stem cells from bone marrow and adipose tissue have gradually been employed to cure animal ailments worldwide (Marx et al. 2014). In a study in pigs with induced proctitis, repeated injections of MSCs were found to modify VEGF expression and its receptor and angiopoietins and fibroblast growth factor-2 (Linard et al. 2013). Furthermore, considerable research in animal models has demonstrated that MSC therapy improves wound healing in goats (Pratheesh et al. 2017), sheep (Martinello et al. 2018), horses (Textor et al. 2018), and dogs (Johnson et al. 2017).

2.8 Conclusions

Stem cells have been the topic of much excitement and controversy among scientists and the general population, and in veterinary regenerative medicine, it is an active research area. The foundations of stem cell research lie not with the famous human embryonic stem cells, but with hematopoietic stem cells (HSCs). The field of stem cell therapy is rapidly evolving, and many clinical trials have been introduced to investigate the use of stem cells in the treatment of a wide range of diseases—the natural function of stem cells in the regeneration of aged or damaged tissue. MSCs are also being explored for use in 3D printing applications because of their unique capacity to form structural tissues. It is hoped that by combining stem cell technologies with different scaffolds able to deliver combinations of growth factors, we may be able to treat conditions in regenerative medicine. MSC therapies have yielded notable results, particularly in treating orthopedic conditions; also, many advancements have been made in treating other conditions such as wound healing. Many studies' positive results indicate great promise for the future of stem cell therapies for various animal diseases, but numerous issues must be addressed. The first one is the autologous MSCs production, and these cells are now used in many clinical trials and treatments; however, this procedure has drawbacks such as the time required to obtain an adequate number of cells from elderly or frail patients or the difficulty of growing MSCs in vitro from patients with various pathologies. As a result, cryopreservation of cells has been widely used to allow for delayed treatment or allogeneic donors, although cryopreservation is not an innocuous process for cells.

On the other hand, producing enough MSCs through in vitro expansion to obtain a clinical dose may influence the MSCs' native properties. Moreover, future research will be needed to determine the effect of age and possibly sex on the therapeutic capabilities of MSC. Furthermore, the entrapment of MSC in the lungs following

Intravenous (IV) administration is another critical issue in systemic stem cell therapies. Although other routes of administration have been considered to avoid lung entrapment, systemic administration of ECV may be the alternative to IV administration of MSCs (Fischer et al. 2009). Multiple parameters, including tissue origin, cryopreservation procedure, culture time and media supplementation with different growth factors, optimal dosage, and in vivo cell delivery, can affect the therapeutic properties of MSCs. Therefore, a deeper understanding of these cell processes would improve the therapeutic outcomes of MSCs. As a result, developing universal protocols for MSC maintenance, banking, and culture would be beneficial. There is still a significant gap between laboratory research and approved stem cell-based products. Future research should develop faster and more reproducible methods for isolating, expanding, and characterizing progenitors from adipose tissue or bone marrow and identifying pathways that activate endogenous stromal cells in the joint. Cryopreservation has intriguing clinical benefits and is required for MSC banking, but its effects on MSC biology are debatable. Hence, further research into improving cryopreservation conditions to ensure the intrinsic biological properties of MSCs is required to extend the utility of MSC banking for subsequent cell therapy. Clinical trials must be completed to assure the safety and efficacy of stem cell therapy. Overall, stem cell-based therapy offers excellent therapeutic potential and may represent a great hope for multiple diseases and degenerative conditions, for which conventional treatments have various side effects and limitations. In the future, advances in pluripotent stem cell techniques will revolutionize livestock breeding.

Nonetheless, significant progress has been made in developing safe and effective stem cell therapies. However, continuous scientific research is undoubtedly needed to fully understand the complexity and severity of specific diseases and the regenerative effects of stem cells. To summarize, many of these research fields attempting to improve stem cell efficiency are ongoing and yielding promising preclinical results, though the translation of their findings into clinical practice appears to be remote.

References

Abdelaz P et al (2019) Reparative dentin formation using stem cell therapy versus calcium hydroxide in direct pulp capping: an animal study. Braz Dent J 30:542–549

Aithal AP, Bairy LK, Seetharam RN (2021) Safety and therapeutic potential of human bone marrow-derived mesenchymal stromal cells in regenerative medicine. Stem Cell Investig 8:8

Ansari M, Amarpal A, Sharma GT et al (2013) Therapeutic potential of canine bone marrow derived mesenchymal stem cells and its conditioned media in diabetic rat wound healing. J Stem Cell Res Ther 3:141

Baiguera S et al (2014) Electrospun gelatin scaffolds incorporating rat decellularized brain extracellular matrix for neural tissue engineering. Biomaterials 35(4):1205–1214

Baird A et al (2018) Osteoblast differentiation of equine induced pluripotent stem cells. Biol Open 7(5):bio033514

Baligar P et al (2017) Bone marrow stem cell therapy partially ameliorates pathological consequences in livers of mice expressing mutant human α1-antitrypsin. Hepatology 65(4): 1319–1335

Banas A et al (2008) IFATS collection: in vivo therapeutic potential of human adipose tissue mesenchymal stem cells after transplantation into mice with liver injury. Stem Cells 26(10): 2705–2712

Barky AR, Ali EMM, Mohamed TM (2017) Stem cells, classifications and their clinical applications. Am J Pharmacol Ther 1:1–7

Barussi FC et al (2016) Intratracheal therapy with autologous bone marrow-derived mononuclear cells reduces airway inflammation in horses with recurrent airway obstruction. Respir Physiol Neurobiol 232:35–42

Bavin EP et al (2015) Equine induced pluripotent stem cells have a reduced tendon differentiation capacity compared to embryonic stem cells. Front Vet Sci 2:55

Bhang SH et al (2012) Transplantation of cord blood mesenchymal stem cells as spheroids enhances vascularization. Tissue Eng A 18(19–20):2138–2147

Bharti MK et al (2020) Effect of cryopreservation on therapeutic potential of canine bone marrow derived mesenchymal stem cells augmented mesh scaffold for wound healing in Guinea pig. Biomed Pharmacother 121:109573

Bhat IA et al (2019) An allogenic therapeutic strategy for canine spinal cord injury using mesenchymal stem cells. J Cell Physiol 234(3):2705–2718

Bigham-Sadegh A et al (2012) Effects of adipose tissue stem cell concurrent with greater omentum on experimental long-bone healing in dog. Connect Tissue Res 53(4):334–342

Bist D et al (2021) Evaluation of canine bone marrow-derived mesenchymal stem cells for experimental full-thickness cutaneous wounds in a diabetic rat model. Expert Opin Biol Ther 21(12):1655–1664

Bliley JM et al (2016) Administration of adipose-derived stem cells enhances vascularity, induces collagen deposition, and dermal adipogenesis in burn wounds. Burns 42(6):1212–1222

Bogliotti YS et al (2018) Efficient derivation of stable primed pluripotent embryonic stem cells from bovine blastocysts. Proc Natl Acad Sci U S A 115(9):2090–2095

Bouchez G et al (2008) Partial recovery of dopaminergic pathway after graft of adult mesenchymal stem cells in a rat model of Parkinson's disease. Neurochem Int 52(7):1332–1342

Canham MA et al (2010) Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. PLoS Biol 8(5):e1000379

Caplan AI (2015) Adult mesenchymal stem cells: when, where, and how. Stem Cells Int 2015:1

Chan B, Leong K (2008) Scaffolding in tissue engineering: general approaches and tissue-specific considerations. Eur Spine J 17(4):467–479

Chandra V, Mankuzhy P, Sharma G (2022) Mesenchymal stem cells in veterinary regenerative therapy: basic physiology to clinical applications. Curr Stem Cell Res Ther 17(3):237–251

Chaudhuri O et al (2020) Effects of extracellular matrix viscoelasticity on cellular behaviour. Nature 584(7822):535–546

Chen L et al (2008) Paracrine factors of mesenchymal stem cells recruit macrophages and endothelial lineage cells and enhance wound healing. PLoS One 3(4):e1886

Chhabra P, Brayman KL (2009) The use of stem cells in kidney disease. Curr Opin Organ Transplant 14(1):72–78

Çolpak H et al (2019) Vertical ridge augmentation using guided bone regeneration procedure and dental pulp derived mesenchymal stem cells with simultaneous dental implant placement: a histologic study in a sheep model. J Stomatol, Oral Maxillofac Surg 120(3):216–223

Dabrowska S et al (2021) Immunomodulatory and regenerative effects of mesenchymal stem cells and extracellular vesicles: therapeutic outlook for inflammatory and degenerative diseases. Front Immunol 11:3809

Dai L-J et al (2009) The therapeutic potential of bone marrow-derived mesenchymal stem cells on hepatic cirrhosis. Stem Cell Res 2(1):16–25

Dakin S et al (2011) The relationship between in vivo limb and in vitro tendon mechanics after injury: a potential novel clinical tool for monitoring tendon repair. Am J Vet Res 43(4):418–423
Davis AB, Schnabel LV, Gilger BC (2019) Subconjunctival bone marrow-derived mesenchymal stem cell therapy as a novel treatment alternative for equine immune-mediated keratitis: a case series. Vet Ophthalmol 22(5):674–682
Deng J et al (2021) 3D spheroids of human placenta-derived mesenchymal stem cells attenuate spinal cord injury in mice. Cell Death Dis 12(12):1–15
Devireddy LR et al (2017) Questions and challenges in the development of mesenchymal stromal/stem cell-based therapies in veterinary medicine. Tissue Eng Part B Rev 23(5):462–470
Djouad F et al (2009) Mesenchymal stem cells: innovative therapeutic tools for rheumatic diseases. Nat Rev Rheumatol 5(7):392–399
Dominici M et al (2006) Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. Cytotherapy 8(4):315–317
Dubey NK et al (2018) Combating osteoarthritis through stem cell therapies by rejuvenating cartilage: a review. Stem Cells Int 2018:1
Dubey S et al (2021) 3-D macro/microporous-nanofibrous bacterial cellulose scaffolds seeded with BMP-2 preconditioned mesenchymal stem cells exhibit remarkable potential for bone tissue engineering. Int J Biol Macromol 167:934–946
Dubey S, Shivahare R, Sharma GT (2022) Nanomaterials, neural stem cells, and the path to neural tissue engineering. In: Engineered nanomaterials for innovative therapies and biomedicine. Springer, pp 99–141
El-Menoufy H et al (2010) The role of bone marrow-derived mesenchymal stem cells in treating formocresol induced oral ulcers in dogs. J Oral Pathol Med 39(4):281–289
El-Tookhy OS et al (2017) Histological evaluation of experimentally induced critical size defect skin wounds using exosomal solution of mesenchymal stem cells derived microvesicles. Int J Stem Cells 10(2):144
Filippi M et al (2020) Natural polymeric scaffolds in bone regeneration. Front Bioeng Biotechnol 8:474
Fischer UM et al (2009) Pulmonary passage is a major obstacle for intravenous stem cell delivery: the pulmonary first-pass effect. Stem Cells Dev 18(5):683–692
Friedenstein A, Piatetzky-Shapiro I, Petrakova K (1966) Osteogenesis in transplants of bone marrow cells. J Embryol Exp Morpholog 16(3):381–390
Frisbie DD, Stewart MC (2011) Cell-based therapies for equine joint disease. Vet Clin North Am Equine Pract 27(2):335–349
Gattazzo F, Urciuolo A, Bonaldo P (2014) Extracellular matrix: a dynamic microenvironment for stem cell niche. Biochim Biophys Acta 1840(8):2506–2519
Goel S et al (2011) Spermatogonial stem cells in the testis of an endangered bovid: Indian black buck (Antilope cervicapra L). Anim Reprod Sci 126(3–4):251–257
Gomiero C et al (2018) Allogeneic mesenchymal stem cells improve the wound healing process of sheep skin. BMC Vet Res 14(1):202
Guan K et al (2006) Pluripotency of spermatogonial stem cells from adult mouse testis. Nature 440(7088):1199–1203
Han X et al (2011) Generation of induced pluripotent stem cells from bovine embryonic fibroblast cells. Cell Res 21(10):1509–1512
Harding C, Heuser J, Stahl P (1983) Receptor-mediated endocytosis of transferrin and recycling of the transferrin receptor in rat reticulocytes. J Cell Biol 97(2):329–339
Hass R et al (2011) Different populations and sources of human mesenchymal stem cells (MSC): a comparison of adult and neonatal tissue-derived MSC. Cell Commun Signal 9(1):1–14
Heldring N et al (2015) Therapeutic potential of multipotent mesenchymal stromal cells and their extracellular vesicles. Hum Gene Ther 26(8):506–517
Hinderer S, Layland SL, Schenke-Layland K (2016) ECM and ECM-like materials—biomaterials for applications in regenerative medicine and cancer therapy. Adv Drug Deliv Rev 97:260–269

Hou D-R et al (2016) Derivation of porcine embryonic stem-like cells from in vitro-produced blastocyst-stage embryos. Sci Rep 6(1):1–13
Huang YZ et al (2020) Mesenchymal stem cells for chronic wound healing: current status of preclinical and clinical studies. Tissue Eng Part B Rev 26(6):555–570
Hwang NS et al (2009) Mesenchymal stem cell differentiation and roles in regenerative medicine. Wiley Interdiscip Rev Syst Biol Med 1(1):97–106
Hynes RO (2009) The extracellular matrix: not just pretty fibrils. Science 326(5957):1216–1219
Ikehara S (2010) The future of stem cell transplantation in autoimmune disease. Clin Rev Allergy Immunol 38(2):292–297
Jayaram P et al (2019) Bone marrow–derived and adipose-derived mesenchymal stem cell therapy in primary knee osteoarthritis: a narrative review. PM R 11(2):177–191
Jeyaraman M et al (2021) Mesenchymal stem cell–derived exosomes: a potential therapeutic avenue in knee osteoarthritis. Cartilage 13(1_suppl):1572S–1585S. https://doi.org/10.1177/1947603520962567
Johnson V et al (2017) Activated mesenchymal stem cells interact with antibiotics and host innate immune responses to control chronic bacterial infections. Sci Rep 7(1):1–18
Jung D-I et al (2009) A comparison of autologous and allogenic bone marrow-derived mesenchymal stem cell transplantation in canine spinal cord injury. J Neurol Sci 285(1–2):67–77
Keefer CL (2004) Production of bioproducts through the use of transgenic animal models. Anim Reprod Sci 82:5–12
Kim E et al (2019a) Neural induction of porcine-induced pluripotent stem cells and further differentiation using glioblastoma-cultured medium. J Cell Mol Med 23(3):2052–2063
Kim SE et al (2019b) Intra-articular umbilical cord derived mesenchymal stem cell therapy for chronic elbow osteoarthritis in dogs: a double-blinded, placebo-controlled clinical trial. Front Vet Sci 6:474
Kirkby Shaw K et al (2020) Fundamental principles of rehabilitation and musculoskeletal tissue healing. Vet Surg 49(1):22–32
Kitada M, Dezawa M (2009) Induction system of neural and muscle lineage cells from bone marrow stromal cells; a new strategy for tissue reconstruction in degenerative diseases. Histol Histopathol 24(5):631–642
Kumar D et al (2021) Perspectives of pluripotent stem cells in livestock. World J Stem Cells 13(1):1
Langhans SA (2018) Three-dimensional in vitro cell culture models in drug discovery and drug repositioning. Front Pharmacol 9:6
Linard C et al (2013) Repeated autologous bone marrow-derived mesenchymal stem cell injections improve radiation-induced proctitis in pigs. Stem Cells Transl Med 2(11):916–927
Liu H et al (2010) Cell therapy of congenital corneal diseases with umbilical mesenchymal stem cells: lumican null mice. PLoS One 5(5):e10707
Liu Z et al (2018) Human umbilical cord mesenchymal stem cells improve irradiation-induced skin ulcers healing of rat models. Biomed Pharmacother 101:729–736
Luo J et al (2017) Vascular smooth muscle cells derived from inbred swine induced pluripotent stem cells for vascular tissue engineering. Biomaterials 147:116–132
Luria E et al (1987) Bone formation in organ cultures of bone marrow. Cell Tissue Res 248(2):449–454
Magri C et al (2019) Comparison of efficacy and safety of single versus repeated intra-articular injection of allogeneic neonatal mesenchymal stem cells for treatment of osteoarthritis of the metacarpophalangeal/metatarsophalangeal joint in horses: a clinical pilot study. PLoS One 14(8):e0221317
Mahendru D et al (2021) Neuroprotective effect of bone marrow-derived mesenchymal stem cell secretome in 6-OHDA-induced Parkinson's disease. Regen Med 16(10):915–930
Mankuzhy PD et al (2021) The preclinical and clinical implications of fetal adnexa derived mesenchymal stromal cells in wound healing therapy. Wound Repair Regen 29:347
Martin GR (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A 78(12):7634–7638

Martinello T et al (2018) Allogeneic mesenchymal stem cells improve the wound healing process of sheep skin. BMC Vet Res 14(1):1–9
Marx C et al (2014) Acupoint injection of autologous stromal vascular fraction and allogeneic adipose-derived stem cells to treat hip dysplasia in dogs. Stem Cells Int 2014:1
Min B-H et al (2007) The fate of implanted autologous chondrocytes in regenerated articular cartilage. Proc Inst Mech Eng H J Eng Med 221(5):461–465
Mueller-Klieser W (1987) Multicellular spheroids. J Cancer Res Clin Oncol 113(2):101–122
Muniswami DM et al (2018) Motor recovery after transplantation of bone marrow mesenchymal stem cells in rat models of spinal cord injury. Ann Neurosci 25(3):126–140
Murphy SV et al (2020) Amnion membrane hydrogel and amnion membrane powder accelerate wound healing in a full thickness porcine skin wound model. Stem Cells Transl Med 9(1):80–92
Nagano H et al (2021) Enhanced cellular engraftment of adipose-derived mesenchymal stem cell spheroids by using nanosheets as scaffolds. Sci Rep 11(1):1–12
Niemann H, Kues WA (2003) Application of transgenesis in livestock for agriculture and biomedicine. Anim Reprod Sci 79(3–4):291–317
Niemann H, Kues WA (2007) Transgenic farm animals: an update. Reprod Fertil Dev 19(6): 762–770
Nitahara-Kasahara Y et al (2012) Long-term engraftment of multipotent mesenchymal stromal cells that differentiate to form myogenic cells in dogs with Duchenne muscular dystrophy. Mol Ther 20(1):168–177
Ochiai H et al (2017) Transplanted mesenchymal stem cells are effective for skin regeneration in acute cutaneous wounds of pigs. Regen Ther 7:8–16
Oreffo RO et al (2005) Mesenchymal stem cells. Stem Cell Rev 1(2):169–178
Papadimitriou L et al (2020) Biofabrication for neural tissue engineering applications. Mater Today Bio 6:100043
Park I-S, Chung P-S, Ahn JC (2015) Enhancement of ischemic wound healing by spheroid grafting of human adipose-derived stem cells treated with low-level light irradiation. PLoS One 10(6): e0122776
Peer BA et al (2022) Comparative evaluation of fracture healing potential of differentiated and undifferentiated Guinea pig and canine bone marrow-derived mesenchymal stem cells in a Guinea pig model. Tissue Cell 76:101768
Peralta O et al (2020) Safety and efficacy of a mesenchymal stem cell intramammary therapy in dairy cows with experimentally induced Staphylococcus aureus clinical mastitis. Sci Rep 10(1): 1–12
Petrov R et al (2003) Influence of topically applied cold treatment on core temperature and cell viability in equine superficial digital flexor tendons. Am J Vet Res 64(7):835–844
Pittenger MF et al (1999) Multilineage potential of adult human mesenchymal stem cells. Science 284(5411):143–147
Popp FC et al (2009) Mesenchymal stem cells can affect solid organ allograft survival. Transplantation 87(9S):S57–S62
Prabhakar S et al (2012) Autologous bone marrow-derived stem cells in amyotrophic lateral sclerosis: a pilot study. Neurol India 60(5):465
Pratheesh M et al (2017) Comparative study on characterization and wound healing potential of goat (Capra hircus) mesenchymal stem cells derived from fetal origin amniotic fluid and adult bone marrow. Res Vet Sci 112:81–88
Prządka P et al (2021) The role of mesenchymal stem cells (MSCs) in veterinary medicine and their use in musculoskeletal disorders. Biomol Ther 11(8):1141
Quimby JM et al (2011) Evaluation of intrarenal mesenchymal stem cell injection for treatment of chronic kidney disease in cats: a pilot study. J Feline Med Surg 13(6):418–426
Quimby JM et al (2013) Safety and efficacy of intravenous infusion of allogeneic cryopreserved mesenchymal stem cells for treatment of chronic kidney disease in cats: results of three sequential pilot studies. Stem Cell Res Ther 4(2):1–12

Raghav A et al (2021) Mesenchymal stem cell-derived exosomes exhibit promising potential for treating SARS-CoV-2-infected patients. Cell 10(3):587
Ryu H-H et al (2012) Comparison of mesenchymal stem cells derived from fat, bone marrow, Wharton's jelly, and umbilical cord blood for treating spinal cord injuries in dogs. J Vet Med Sci 74:12–0065
Saha K et al (2008) Substrate modulus directs neural stem cell behavior. Biophys J 95(9): 4426–4438
Santos JM et al (2015) Three-dimensional spheroid cell culture of umbilical cord tissue-derived mesenchymal stromal cells leads to enhanced paracrine induction of wound healing. Stem Cell Res Ther 6(1):1–19
Satoh H et al (2004) Transplanted mesenchymal stem cells are effective for skin regeneration in acute cutaneous wounds. Cell Transplant 13(4):405–412
Scott L et al (2021) Carbon nanotube-based scaffolds for cardiac tissue engineering—systematic review and narrative synthesis. Bioengineering 8(6):80
Seil JT, Webster TJ (2010) Electrically active nanomaterials as improved neural tissue regeneration scaffolds. Wiley Interdiscip Rev Nanomed Nanobiotechnol 2(6):635–647
Sgrignoli MR et al (2019) Reduction in the inflammatory markers CD4, IL-1, IL-6 and TNFα in dogs with keratoconjunctivitis sicca treated topically with mesenchymal stem cells. Stem Cell Res 39:101525
Shabir U et al (2022) Smad4 and γ-secretase knock-down effect on osteogenic differentiation mediated via Runx2 in canine mesenchymal stem cells. Res Vet Sci 145:116–124
Sharma R et al (2014) Generation of functional neurons from feeder-free, keratinocyte-derived equine induced pluripotent stem cells. Stem Cells Dev 23(13):1524–1534
Sharun K et al (2020) Clinical evaluation following the percutaneous transplantation of allogenic bone marrow-derived mesenchymal stem cells (aBM-MSC) in dogs affected by vertebral compression fracture. Vet Anim Sci 10:100152
Shi H et al (2007) Nanomed.: Nanotechnol. Biol Med 3:266–272
Shi H et al (2018) Improved biphasic calcium phosphate combined with periodontal ligament stem cells may serve as a promising method for periodontal regeneration. Am J Transl Res 10(12): 4030
Shrestha C et al (2013) Enhanced healing of diabetic wounds by subcutaneous administration of human umbilical cord derived stem cells and their conditioned media. Int J Endocrinol 2013: 2013:592454
Smith R et al (2003) Isolation and implantation of autologous equine mesenchymal stem cells from bone marrow into the superficial digital flexor tendon as a potential novel treatment. Equine Vet J 35(1):99–102
Smith RKW et al (2013) Beneficial effects of autologous bone marrow-derived mesenchymal stem cells in naturally occurring tendinopathy. PLoS One 8(9):e75697
Strasser H et al (2006) Stem cell therapy for urinary stress incontinence. J Urol 175(5):1801–1801
Sun Z et al (2013) Scaffold-based delivery of autologous mesenchymal stem cells for mandibular distraction osteogenesis: preliminary studies in a porcine model. PLoS One 8(9):e74672
Taib NA, Rahman MR et al (2022) A review on poly lactic acid (PLA) as a biodegradable polymer. Polym Bull 80(2):1179–1213
Textor JA et al (2018) Allogeneic stem cells alter gene expression and improve healing of distal limb wounds in horses. Stem Cells Transl Med 7(1):98–108
Thomson JA et al (1995) Isolation of a primate embryonic stem cell line. Proc Natl Acad Sci U S A 92(17):7844–7848
Thomson JA et al (1998) Embryonic stem cell lines derived from human blastocysts. Science 282(5391):1145–1147
Tortorella S et al (2022) Laser assisted bioprinting of laminin on biodegradable PLGA substrates: effect on neural stem cell adhesion and differentiation. Bioprinting 26:e00194
Udalamaththa VL et al (2020) Therapeutic uses of post-partum tissue-derived mesenchymal stromal cell secretome. Indian J Med Res 152(6):541

Van Loon VJ et al (2014) Clinical follow-up of horses treated with allogeneic equine mesenchymal stem cells derived from umbilical cord blood for different tendon and ligament disorders. Vet Q 34(2):92–97
Vanikar A et al (2010) Cotransplantation of adipose tissue-derived insulin-secreting mesenchymal stem cells and hematopoietic stem cells: a novel therapy for insulin-dependent diabetes mellitus. Stem Cells Int 2010:582382
Venkataiah VS et al (2019) Periodontal regeneration by allogeneic transplantation of adipose tissue derived multi-lineage progenitor stem cells in vivo. Sci Rep 9(1):1–12
Vijayavenkataraman S (2020) Nerve guide conduits for peripheral nerve injury repair: a review on design, materials and fabrication methods. Acta Biomater 106:54–69
Villarroya-Beltri C et al (2014) Sorting it out: regulation of exosome loading. In: Seminars in cancer biology, vol 28. Elsevier, p 3
Viswanathan P et al (2015) 3D surface topology guides stem cell adhesion and differentiation. Biomaterials 52:140–147
Wei R et al (2020) Derivation of endothelial cells from porcine induced pluripotent stem cells by optimized single layer culture system. J Vet Sci 21(1):e9
Whitworth DJ et al (2014) derivation of mesenchymal stromal cells from canine induced pluripotent stem cells by inhibition of the TGFβ/activin signaling pathway. Stem Cells Dev 23(24): 3021–3033
Wilmut I et al (1997) Viable offspring derived from fetal and adult mammalian cells. Cloning Stem Cells 385(6619):810–813
Wu Y et al (2007) Mesenchymal stem cells enhance wound healing through differentiation and angiogenesis. Stem Cells 25(10):2648–2659
Xue J, Pisignano D, Xia Y (2020) Maneuvering the migration and differentiation of stem cells with electrospun nanofibers. Adv Sci 7(15):2000735
Zhu S et al (2011) Effects of intrahepatic bone-derived mesenchymal stem cells autotransplantation on the diabetic beagle dogs. J Surg Res 168(2):213–223
Zhu R et al (2019) Electrical stimulation affects neural stem cell fate and function in vitro. Exp Neurol 319:112963

Chapter 3
Role of Proteomics in Understanding Livestock Health

Shalini Jaswal and Sudarshan Kumar

Abstract Proteomics provides the most effective platform for disease diagnosis and monitoring. Proteomic analysis is a multistep procedure starting from protein isolation from the biological sample, trypsin digestion, peptide fractionation, nanoLC-MS/MS analysis, RAW data analysis for identification of protein, and bioinformatics analysis to study the biological significance of identified proteins. Although proteomics has contributed immensely to the field of human health research, limited studies have been reported on its application in animal research due to its high cost and limited progress in veterinary research. In this chapter, we are summarizing applications of proteomics in understanding the molecular basis of diseases of livestock animals, including bovine mastitis.

Keywords Proteomics · Disease diagnosis · Mastitis · Peptide fractionation · nanoLC-MS/MS analysis

3.1 Introduction

A number of animals are classified as livestock species, including cows, buffalos, sheep, goats, horses, and pigs. They contribute immensely towards socioeconomic development. These domesticated animals are reared to meet a variety of human needs, such as labor, food, wool, and leather. Infectious illnesses are the most serious hazard to cattle. As a result, animals must be monitored on a regular basis for the emergence of such diseases in order to maintain their health. There is also a need for better strategies to improve animal health and control decrease epidemics.

Proteomics provides the most effective platform for disease diagnosis and monitoring. It refers to the study of proteins present in a given biological system, including tissues, cells, and body fluids, like urine, blood, saliva, tear etc. Proteomic analysis is a multistep procedure (Fig. 3.1) that entails protein isolation from the

S. Jaswal · S. Kumar (✉)
Animal Biotechnology Center, ICAR-National Dairy Research Institute, Karnal, Haryana, India

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_3

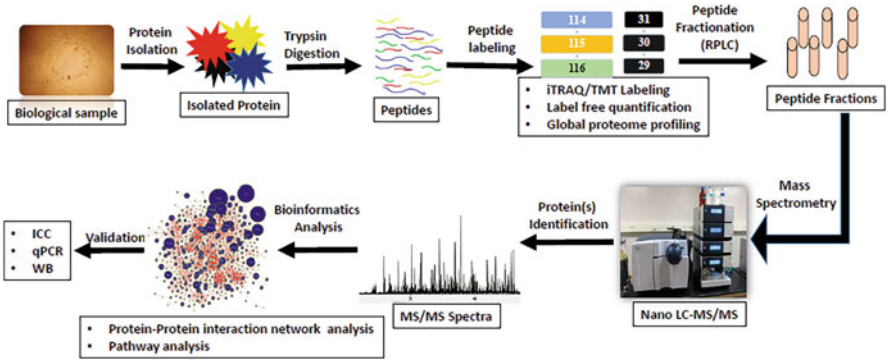

Fig. 3.1 Schematic diagram depicting the workflow for LC-MS-based proteomics analysis of biological samples

biological sample, trypsin digestion, peptide fractionation, nanoLC-MS/MS analysis, RAW data analysis for identification of protein, and Bioinformatics analysis to study the biological significance of identified proteins. Depending upon the aim of the study, proteomics analysis can be used for either global profiling or differential proteome analysis. Global profiling is used to characterize the protein profile of the biological system under investigation, whereas differential proteome analysis is used to investigate and compare the proteome profiles of samples under various conditions and treatments. Label-free or labeling techniques can be used for differential proteome analysis. The labeling approaches are based on TMT and iTRAQ, which enable multiplexing (Chandramouli and Qian 2009).

Proteomics has contributed immensely to the field of human health research. However, to date, limited studies have been reported on its application in animal research due to its high cost and limited progress in veterinary research. The altered protein profile helps to understand the molecular basis of various pathological states, infections, and stressful conditions in animals. Also, the proteomics approach helps in finding the biomarkers for pathophysiological alterations . The biomarker study involves two main phases including discovery and validation. The discovery phase refers to the identification of proteins in a particular sample, where we can quantify the expression of a particular protein in normal and diseased or altered states. However, targeted proteomics is required to establish its role as a biomarker. The uncharacterized animal genomes limit the application of proteomics approaches for the study of animals in various aspects. In-depth proteome profiling of BuMECs identified a total of 12,609 proteins, which can be used as dataset for proteome-based genome annotations (Jaswal et al. 2020).

Animal proteomics is a growing field that provides a great platform for understanding the pathogenesis of various origins in domestic animals. Proteomics used advanced analytical methods such as liquid chromatography and mass spectrometry (M.S.) to characterize the protein composition of biological samples. The rapid advancement in the proteomics approaches has enabled us to discover and

understand the complexity of the protein machinery of cells, which governs various physiological functions. Targeted proteomics helps in detection and quantitation, and hence validation, of potential biomarkers associated with a physiological state. Such biomarkers play an essential role in the diagnosis and monitoring of disease, hence its management. Unlike in humans, a limited number of studies have been done so far on animal health and pathogenesis. This chapter focuses deeply on the application of proteomics in understanding the molecular basis of diseases associated with livestock species.

3.2 Proteomics Application in Sheep Health

The application of proteomics in the treatment of sheep diseases has been limited. However, many proteomics-based studies have been done to understand the pathogenesis of many diseases in sheep. The proteome profile of serum, acellular proteome, helps in the analysis and interpretation of the physiological changes governing the normal and diseased state in an animal. The two-dimensional gel electrophoresis (2-DE) followed by M.S. analysis of 250 gel spots identified a total of 42 proteins. The expression of transthyretin was significantly downregulated. However, endorphin and alpha1β glycoprotein were found to be highly upregulated. Many of the proteins belonged to the clinical biomarkers and were common among other species too. The identified proteins are the potential biomarkers for the detection and monitoring of physiology as well as metabolism in ovines (Chiaradia et al. 2012). Another study used serum proteome profiling to examine changes in serum proteins in animals in the peripartum period and those with moderate respiratory illness. It would aid in the betterment of sheep clinical care. The differential proteome analysis revealed 17 proteins that were differentially expressed across the two states, with fold changes of >2 and 0.05 for upregulated and downregulated proteins, respectively. A total of 379 proteins were discovered using MS. Their association with 127 signaling pathways was discovered by bioinformatics study (Chemonges et al. 2016).

The fat deposit in the sheep tail acts as a source of energy during adverse situations. However, the molecular process underlying adipose tissue's physiological significance in the tail is unknown. In Kazakh, Lanzhou, Hu, Merino, and Tibetan tail adipose tissue, M.S. revealed a total of 2303, 2137, 2666, 1932, and 1941 proteins, respectively. Merino and Tibetan have thin fat tissue, whereas Kazakh, Lanzhou, and Hu have fat tail tissue. Overall, 3400 proteins were discovered, with 1209 of these being shared by all five breeds. According to label-free quantification, 804 proteins were differentially expressed across the three fat-tailed lambs and the two thin-tailed lambs, with FABP4 having the greatest expression across all five breeds. Furthermore, enrichment analysis revealed that these proteins are involved in a variety of biological processes, including metabolism, the PPAR pathway, and adaptive thermogenesis. One hundred and seventy DEPs were linked to carboxylic and metabolic processes, oxidation-reduction processes, and small

molecule metabolic processes, among others. Several proteins have been linked to the elongation, metabolism, and breakdown of fatty acids. The RT-qPCR analysis suggested that the expression pattern of 24 proteins was consistent with their abundance at the protein level. Also, immunofluorescence and immunohistochemistry analyzed the expression of FAB4 protein in the capillary wall of the fat tail, which suggested its role in the transport of fatty acids from the blood to the adipocytes. The study helped to understand the difference in protein composition of tail fat tissue among the fat and thin tails in lambs. The study ensures the crucial role of FAB4 protein in the tail and is a crucial target to be considered as the biomarker for the fat-tailed feature in sheep, which needs further study. Besides FABP4, other proteins, including ACSL1, ACACA, FASN, HSD17b4, and ACLY were significantly enriched in the fatty-tailed tissue. The study provides new insight into the molecular significance of tailed fat tissue, which may have clinical importance too; however, it needs further study (Han et al. 2021).

The mammary glands of sheep ($n = 5$) were injected with Mannheimia hemolytica to investigate the molecular basis of mastitis. Mastitis was confirmed in the mammary gland utilizing clinical, cytological, microbiological, and histological tests. After 12 h of infection, a proteomics investigation was undertaken. Following two-dimensional gel electrophoresis using IPG strips (18 cm and P.I. 3–10) and non-gradient 12 percent SDS-gel, MALDI-TOF-based M.S. analysis of protein gel spots identified a total of 33, 89, and 20 differentially expressed proteins in the blood plasma, whey from the inoculated side of the udder, and whey from the contralateral side of the udder, respectively. In contrast, 79 whey proteins were expressed differentially in the infected and contralateral glands (Katsafadou et al. 2019).

The fluid, known as tears or tear film, covers the exterior surface of the eye and functions as a protective mechanism. Lipids, carbohydrates, proteins, and electrolytes are among the components of this fluid. Protein expression changes with ocular and systemic illnesses. A proteomic technique was utilized to investigate species-specific differences in ocular fluid. Tears were taken from 25 sheep, 50 cows, and 50 camels, as well as 25 humans. Protein isolation was performed on the samples, and SDS-PAGE analysis of an equal amount of protein from each species was performed. In the tears of cows, humans, camels, and sheep, a total of d 13, 12, 21, and 12 unique bands were discovered, respectively.

In addition, 2-D gel runs revealed 182, 223, 217, and 241 protein spots in human, cow, sheep, and camel samples, respectively. Furthermore, MS analysis revealed that numerous proteins such as lysozyme, immunoglobulin, lactoferrin, and serum albumin were abundant across all species, implying that they play a vital role. Furthermore, numerous proteins displayed species-specific expressions, such as the vitelline membrane outer layer protein 1 (VMO1) protein, which was discovered in camel and sheep tears but not in human or cow tears. Lipocalins were also found only in the tears of sheep and cows. Variations in the proteome expression profile could be caused by evolution or environmental influences. The presence of such camel-specific proteins may aid in understanding their significance in camel survival in harsh desert circumstances such as high humidity, high temperature, and dust.

This could help researchers better understand the biology and therapy of dry eye disease in humans (Shamsi et al. 2011).

3.3 Proteomics Application in Bovine Health

Mastitis is the most prevalent health problem in dairy cows, followed by fertility problems and lameness (Altena et al. 2016). It reduces milk output, affects milk quality, and deteriorates animal health, resulting in significant economic losses in the dairy industry. Mastitis is the most common infectious disease in dairy cows, costing the worldwide dairy sector €16–26 billion per year (www.dairy.ahdb.org.uk) and €61–97 per animal to farmers.

Mastitis refers to the inflammation of the mammary gland in response to the bacterial infection mainly by *Staphylococcus aureus*. It occurs in two phases including subclinical and clinical mastitis. It can be detected by measuring the number of somatic cells in the milk samples. Its number varies; in normal milk samples its number is 7×10^4 per mL of milk, whereas for subclinical samples, its value is $2–5 \times 10^5$, and for clinical samples, its value is $\geq 5 \times 10^5$. Clinical mastitis is characterized by the morphological changes in the neutrophils that lead to the development of pseudopods on their surface. Also, it causes inflammation and reddening of the mammary gland, loss of appetite, fever, and pain. In response to the pathogen attack during mastitis, neutrophils form aggregates with the pathogens known as extracellular traps (NETs) (Swain et al. 2014). These visible signs make it easy to diagnose mastitis. It deteriorates the milk quality and production and also changes the serum composition. Mastitis induces curdling and precipitation of casein in milk. Because casein accounts for roughly 80% of total milk proteins, whey fraction analysis is favored for studying the less abundant proteins. Mastitis is a cow disease that has become a problem in the dairy industry. Potential biomarkers that can aid in the early diagnosis of mastitis, i.e., subclinical mastitis, are needed. Serum and whey samples from normal healthy and mastitic cows were analyzed using 2-DE and MALDI-TOF MS/MS.

The healthy and mastitic cows' serum samples had similar compositions, according to M.S. analysis. The composition of whey samples, on the other hand, varies, which could be attributed to protein translocation from the blood to the mammary gland. Caseins and other milk proteins were detected in both samples; however, lactotransferrin was only found in the mammary gland with mastitic characteristics. Surprisingly, the mastitic whey sample had a higher number of proteins than the healthy whey sample (Alonso-Fauste et al. 2012). In a similar study, proteomics analysis was conducted to study the proteomic changes in milk from cows having subclinical mastitis. A total of 109 proteins were differentially expressed among the subclinical and clinical mastitic milk samples. Bioinformatics analysis suggested the positive correlation of overexpressed differentially expressed proteins (DEPs) with immune-related pathways, such as in the activation of the host defense system and in the overproduction of immunomodulatory molecules such as

inflammatory molecules and cytokines. A quantitative proteomics study using milk suggested CHI3L1, LBP, GSN, GCLC, and PIGR as the potential diagnostic biomarkers for the detection of subclinical mastitis in bovine (Bathla et al. 2020). A quantitative examination of milk whey proteins derived from cows infected with *Streptococcus uberis* in the mammary gland revealed that acute phase proteins (APPs) and antimicrobial peptides such as cathelicidins and peptidoglycan recognition protein were upregulated. The release of APPs into the blood is caused by the inflammation of the mammary gland during mastitis infection (Mudaliar et al. 2016). On the third day of lactation, the concentration of haptoglobin was found to be considerably higher, suggesting that it could be utilized as a marker for the detection of subclinical mastitis postpartum (Simões et al. 2018).

Previously, several biomarkers have been established to monitor the state of health in cows, such as nonesterified fatty acids (NEFA) for clinical ketosis prepartum and mastitis postpartum. APPs including haptoglobin and serum amyloid A (SAA3) in blood, increased population of somatic cells, and lactate dehydrogenase (LDH) expression in milk have all been identified as markers indicating infection and inflammation of the mammary gland, often known as subclinical mastitis, in cows. Following calving, cows are more vulnerable to illnesses. This is due to insufficient dietary intake, which resulted in a negative energy balance. Milk, interestingly, is a good place to look for disease-resistant biomarkers. Interestingly, milk serves as a great source to investigate disease-resistant biomarkers. To do so, milk from healthy cows ($n = 4$) was compared to milk from poor healthy cows ($n = 4$) at the protein level to find novel disease-resistant biomarkers. A total of 78 proteins were discovered in the NanoLC-MS/M.S. analysis of whey samples, 13 of which were strongly expressed in low-resistant cows compared to high-resistant cows. These proteins have the potential to be used as disease-resistant biomarkers. ELISA was used to confirm the expression of one of these proteins, lactoferrin (L.F.), in resistant and sick cows. Its increased abundance causes lameness in cows, and hence, it can be employed as a prognostic biomarker for early culling (Altena et al. 2016).

Proteomics has great application in understanding the nutritional adaption in cows with different physiological conditions. Dairy cows undergo systemic mammary gland inflammation during the transition from late pregnancy to early lactation. It leads to reduced feed intake and increased concentration of non-esterified fatty acids and ketone bodies. Interestingly, the supplementation of cow feed with essential fatty acids (EFA), including linoleic acid (L.A.) and α-linolenic acid (ALA), as well as conjugated linoleic acids (CLA) during the early lactation period has a positive effect on the reduction of milk fat. To understand the metabolic adaptation while transition, plasma proteomics was performed in the blood sample from cows supplementation of cow feed with EFA (Linseed and Safflower oil) and CLA (Lutalin) at 21 days before childbirth (prepartum) (−21) and at three different time points of postpartum (after childbirth), i.e., +1, +28, and + 63. Cow feed with coconut oil was taken as the control. The MS analysis identified a total of 241 proteins with ≥2 unique peptides and 1% false discovery rate (FDR). Label-free quantitation-based differential proteome analysis suggested that from day −21 to

+1 the expression of 14 proteins increased, whereas 49 proteins decreased. KEGG pathways analysis suggested the involvement of upregulated proteins in cholesterol metabolism and immunity. On the other hand, from day +1 to +28, 29 proteins were upregulated and 10 proteins were downregulated. From day +28 to +63, a single protein was upregulated and 24 proteins were downregulated. Bioinformatics analysis suggested that the supplementation of cow feed with EFA and CLA upregulates the expression of apolipoproteins (APOs) including A1, A4, and C3. These APOs are involved in fat digestion and absorption, cholesterol metabolism, and the PPAR signaling pathway (Veshkini et al. 2022).

Cows' hormone levels, notably progesterone, and metabolic rate vary dramatically during the transition from late pregnancy to early lactation. In differentiated buffalo mammary epithelial cells (BuMECs), the expression of DEPs linked to 46 metabolic pathways was found to be higher (Jaswal et al. 2021). Janjanam et al. (2014) discovered 41 differentially regulated proteins in milk epithelial cells across three lactation stages: early, mid, and late lactation. These DEPs were found to have a role in metabolism, binding, and catalytic activities. A higher diet is required to meet the nutrient requirement for milk synthesis. Cows also go through physiological changes such as enhanced gluconeogenesis, lipid mobilization, and bone resorption.

Adipose tissue mobilization raises the concentration of nonessential fatty acids (NEFAs) in milk, which is then utilized for fat synthesis by the mammary gland or gluconeogenesis by the liver. The liver uses a little part of NEFA, while the majority of the NEFA that isn't used is converted into ketone bodies. Ketosis is caused by a buildup of ketone bodies, which causes lameness, abomasum displacement, inflammation, oxidative damage, and insulin resistance. M.S. analysis of blood serum was used to better understand the alterations in proteome and metabolic profiles during ketosis. Blood samples were collected from 18 healthy cows and 18 cows afflicted with clinical ketosis. A total of 48 metabolites and 30 proteins were found to be differentially expressed in postpartum controls (H.C., BHBA = 0.650.22 mm) and prepartum controls (PHC, BHBA = 0.420.08 mm) blood samples ($N = 10$ each). Differentially accumulated metabolites (DAMs) were linked to amino acid metabolisms, such as proline and arginine, whereas DEPs were linked to phagocytosis and cholesterol metabolism. Overall, the findings revealed that changed metabolites were engaged in a variety of metabolic pathways, including nucleotide, carbohydrate, and amino acid metabolism. DEPs, on the other hand, were mostly found in disease-related pathways. During the transition period, the most elevated metabolites, such as 4-Hydroxy-6-Methylpyran-2-one and cinnamoyglycine, could be used as biomarkers to diagnose and monitor ketosis in cows (Wu et al. 2020).

During the transition period, there are significant physiological changes that are required for lactation to begin. However, it can cause physiological imbalance (P.I.) in dairy cows, making them more susceptible to illnesses, especially during early lactation. Cows were administered a limited diet for 4 days to induce P.I. during early (49 ± 22 days of milking) and mid-lactation (159 ± 39 days of milking) to investigate the molecular basis of P.I. The extent of P.I. is calculated based on the expression levels of plasma BHBA, glucose, and NEFA. During early lactation,

iTRAQ-based proteomic study revealed that a total of 8 proteins were differentially expressed in the liver tissue of physiologically imbalanced and normal cows. Pyruvate carboxylase (P.C.) and extremely long chain-specific acyl-CoenzymeA dehydrogenase (VLCAD) were two of the proteins that were elevated. These enzymes play a role in the production of glucose and the oxidation of long-chain fatty acids. Three proteins, however, were downregulated: UDP-glucose-6-dehydrogenase (UGDH), glycine N-acyltransferase (GLYAT), and mitochondrial isocitrate NADP+-dependent dehydrogenase (ICDHm). They play an important function in energy metabolism. During mid-lactation, however, 15 proteins were differentially expressed among the same groups. Seven were upregulated, whereas eight were downregulated. The elevated proteins were discovered to be involved in fatty acid oxidation and ketone body production. The downregulated proteins have functions in the anti-oxidative pathways. The study suggested ICDHm and P.C. as the promising biomarkers during early lactation and dihydrolipoamide dehydrogenase (DLD) and alcohol dehydrogenase 4 (ADH-4) as potential biomarkers during mid-lactation for P.I. These markers can be used to track P.I. in breastfeeding animals and will aid in correct animal management (Moyes et al. 2013).

Bovine milk exosomes are membranous vesicles, about 50–100 nm in size, secreted by the epithelial cells extracellularly. They are composed of lipids, proteins, and RNAs (mRNA and miRNA). Their role has been reported in intracellular communication and immunomodulation. They function in the packaging and presentation of antigens to the immune cells. LC-MS analysis identified a total of 2107 proteins; of these few proteins adipohilin, xanthine oxidase, lactadherin, and butryophilin were highly abundant. Bioinformatics analysis suggested their role in various immunity pathways such as in natural killer cell-mediated cytotoxicity, NOD-like receptor signaling pathway, complement and coagulation cascade, B and T cell receptor signaling pathways, etc. (Reinhardt et al. 2012).

Summer temperatures have a significant impact on the dairy industry. This is due to the heat stress which diminishes lactation production and negatively impacts animal health. The normal bovine was kept at low temperature conditions maintained with the help of soakers and fans, whereas under heat stress cows were kept in an environment devoid of such conditions from 46 days before expected calving. Label-free quantitation-based LC-MS analysis identified a total of 3270 proteins in bovine liver tissues sampled from the heat-stressed and normal bovines at 2 days postpartum. Of the total identified proteins, 75 were differentially expressed. These DEPs were found to play an important role in mitochondrial dysfunction and oxidative phosphorylation. The expressions of proteins involved in these pathways were lower in the heat-stressed cow liver tissue as compared to the control samples. The same trend in expression was shown by the proteins involved in amino acid metabolism such as propionyl-CoA carboxylase-α and -β chains. Other pathways affected were glucose and fatty acid metabolism; however, the altered proteins didn't show a particular trend of expression (Skibiel et al. 2018).

3.4 Proteomics Application in Horse Health

Equine melanocytic neoplasm (EMN) is the most frequent malignant tumor illness in older horses. The differences in proteome profile between normal and pathological feces samples were investigated (different stages of EMN including mid and severe). A total of 5901 proteins were discovered by LC-MS analysis of fecal samples from normal and ill horses ($n = 10$), with 109 being differentially expressed ($p < 0.05$). Surprisingly, 28 DEPs were discovered to have metabolic functions, while the remaining 81 were revealed to be engaged in a variety of other processes such as environmental information processing and cellular operations. DEPs with no expression in the control group were chosen to find prospective disease biomarkers. Fourteen such proteins were found to have either exclusive or extremely high expression in mid-EMN. Few of them were flavoprotein domain-containing protein (PPCDC), diacylglycerol kinase (DGKB), Coesterase domain-containing protein, and beta_elim_lyase domain-containing protein. The FOS like 1, AP-1 transcription factor subunit (FOSL1) was exclusively expressed in the mid-EMN. As a result, it has the potential to be employed as a biomarker for the diagnosis and monitoring of mid-EMN. The four recognized cancer biomarkers DGKB, structural maintenance of chromosome 4 (SMC4), mastermind-like transcriptional coactivator 2 (MAML2), and I.G. domain-containing protein and transglutaminase 2 (Tgm2) were also found as possible EMN biomarkers (Tesena et al. 2022).

Endocrinopathic laminitis is caused by metabolic syndrome and obesity, which is another prevalent condition in horses. There is currently no cure for this condition. To further understand the changes in protein molecules, researchers took samples from normal and damaged horses' cardiac and lamellar tissues (horses injected with insulin for 462.3 h to cause laminitis). In lamellar and cardiac tissues, label-free quantification found a total of 514 and 709 proteins, respectively. In the laminitis tissue, a total of 27 lamellar proteins were significantly changed, with 14 being upregulated and 13 being downregulated. Upregulated proteins such as HSP90, 3 fibrinogen isoforms, and alpha-2-macroglobulin were shown to be involved in ribosomal activities, coagulation cascade, and immunity-related pathways, according to bioinformatics research. Proteins with considerable reductions, on the other hand, have a role in cell-cell contacts and focal adhesion. The research identified critical regulatory molecules and molecular targets for the treatment of laminitis (Campolo et al. 2020).

3.5 Conclusion

MS-based proteomics analysis is a powerful tool for identifying biomarkers linked to a variety of pathophysiological conditions in animals. They play a crucial role in proteome-based genome annotations in animals whose genomes are not annotated. As a result, proteomics is critical in the timely monitoring and diagnosis of illnesses

in animals. In the dairy business, proper management will aid in reducing economic losses due to disease.

References

Alonso-Fauste I, Andrés M, Iturralde M, Lampreave F, Gallart J, Álava MA (2012) Proteomic characterization by 2-DE in bovine serum and whey from healthy and mastitis affected farm animals. J Proteome 75(10):3015–3030

Altena SEC, De Klerk B, Hettinga KA, Van Neerven RJJ, Boeren S, Savelkoul HFJ, Tijhaar EJ (2016) A proteomics-based identification of putative biomarkers for disease in bovine milk. Vet Immunol Immunopathol 174:11–18

Bathla S, Sindhu A, Dang A, Kumar S, Dubey SK, Pattnaik S, Rawat P, Chopra A, Kaushik JK, Mohanty AK (2020) Tandem mass tag (TMT)-based quantitative proteomics reveals potential targets associated with onset of sub-clinical mastitis in cows. Sci Rep 10(1):1–10

Campolo A, Frantz MW, De Laat MA, Hartson SD, Furr MO, Lacombe VA (2020) Differential proteomic expression of equine cardiac and lamellar tissue during insulin-induced laminitis. Front Vet Sci 7:308

Chandramouli K, Qian PY (2009) Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. Hum Genomics Proteomics 2009:239204

Chemonges S, Gupta R, Mills PC, Kopp SR, Sadowski P (2016) Characterisation of the circulating acellular proteome of healthy sheep using LC-MS/MS-based proteomics analysis of serum. Proteome Sci 15(1):1–16

Chiaradia E, Avellini L, Tartaglia M, Gaiti A, Just I, Scoppetta F, Czentnar Z, Pich A (2012) Proteomic evaluation of sheep serum proteins. BMC Vet Res 8(1):1–13

Han J, Guo T, Yue Y, Lu Z, Liu J, Yuan C, Niu C, Yang M, Yang B (2021) Quantitative proteomic analysis identified differentially expressed proteins with tail/rump fat deposition in Chinese thin- and fat-tailed lambs. PLoS One 16(2):e0246279

Janjanam J, Singh S, Jena MK, Varshney N, Kola S, Kumar S, Kaushik JK, Grover S, Dang AK, Mukesh M, Prakash BS (2014) Comparative 2D-DIGE proteomic analysis of bovine mammary epithelial cells during lactation reveals protein signatures for lactation persistency and milk yield. PLoS One 9(8):e102515

Jaswal S, Anand V, Kumar S, Bathla S, Dang AK, Kaushik JK, Mohanty AK (2020) In-depth proteome analysis of more than 12,500 proteins in buffalo mammary epithelial cell line identifies protein signatures for active proliferation and lactation. Sci Rep 10(1):1–13

Jaswal S, Anand V, Ali SA, Jena MK, Kumar S, Kaushik JK, Mohanty AK (2021) TMT based deep proteome analysis of buffalo mammary epithelial cells and identification of novel protein signatures during lactogenic differentiation. FASEB J 35(6):e21621

Katsafadou AI, Tsangaris GT, Anagnostopoulos AK, Billinis C, Barbagianni MS, Vasileiou NG, Spanos SA, Mavrogianni VS, Fthenakis GC (2019) Proteomics data of ovine mastitis associated with Mannheimia haemolytica. Data Brief 25:104259

Moyes KM, Bendixen E, Codrea MC, Ingvartsen KL (2013) Identification of hepatic biomarkers for physiological imbalance of dairy cows in early and mid lactation using proteomic technology. J Dairy Sci 96(6):3599–3610

Mudaliar M, Tassi R, Thomas FC, McNeilly TN, Weidt SK, McLaughlin M, Wilson D, Burchmore R, Herzyk P, Eckersall PD, Zadoks RN (2016) Mastitomics, the integrated omics of bovine milk in an experimental model of streptococcus uberis mastitis: 2. Label-free relative quantitative proteomics. Mol Biosyst 12(9):2748–2761

Reinhardt TA, Lippolis JD, Nonnecke BJ, Sacco RE (2012) Bovine milk exosome proteome. J Proteome 75(5):1486–1492

Shamsi FA, Chen Z, Liang J, Li K, Al-Rajhi AA, Chaudhry IA, Li M, Wu K (2011) Analysis and comparison of proteomic profiles of tear fluid from human, cow, sheep, and camel eyes. Invest Ophthalmol Vis Sci 52(12):9156–9165

Simões PBA, Campbell M, Viora L, Gibbons J, Geraghty TE, Eckersall PD, Zadoks RN (2018) Pilot study into milk haptoglobin as an indicator of udder health in heifers after calving. Res Vet Sci 116:83–87

Skibiel AL, Zachut M, do Amaral BC, Levin Y, Dahl GE (2018) Liver proteomic analysis of postpartum Holstein cows exposed to heat stress or cooling conditions during the dry period. J Dairy Sci 101(1):705–716

Swain DK, Kushwah MS, Kaur M, Patbandha TK, Mohanty AK, Dang AK (2014) Formation of NET, phagocytic activity, surface architecture, apoptosis and expression of toll like receptors 2 and 4 (TLR2 and TLR4) in neutrophils of mastitic cows. Vet Res Commun 38(3):209–219

Tesena P, Kingkaw A, Phaonakrop N, Roytrakul S, Limudomporn P, Vongsangnak W, Kovitvadhi A (2022) Faecal proteomics and functional analysis of equine melanocytic neoplasm in Grey horses. Vet Sci 9(2):94

Veshkini A, Hammon HM, Vogel L, Viala D, Delosière M, Tröscher A, Déjean S, Ceciliani F, Sauerwein H, Bonnet M (2022) Plasma proteomics reveals crosstalk between lipid metabolism and immunity in dairy cows receiving essential fatty acids and conjugated linoleic acid. Sci Rep 12(1):1–12

Wu ZL, Chen SY, Hu S, Jia X, Wang J, Lai SJ (2020) Metabolomic and proteomic profiles associated with ketosis in dairy cows. Front Genet 11:1542

Chapter 4
Microbial Genomics and Antimicrobial Susceptibility Testing

Deepali Kalambhe, Lokesh K.M., Gourab Basak, Sumeet Singh, and Abhilash Jadhao

Abstract Microbial genomics plays an essential role in determining antimicrobial susceptibility. The emergence and rapid spread of antimicrobial resistance against commonly used therapeutic drugs have increased the importance of antimicrobial susceptibility testing in clinically important bacterial infections. The latest molecular methods clinched the interest in detecting resistance genes or allelic mutations. Molecular approaches specifically target nucleic acid amplification preceded by the sequencing of amplified products or whole-genome sequencing. The novel bioinformatics tools and databases on antimicrobial resistance genes have strengthened the understanding of the molecular mechanisms of antibiotic resistance in microorganisms. They have also improved the possibility of accurately predicting antibiotic resistance genes within environmental, agricultural, veterinary, and human health care settings. In contrast, traditional antimicrobial susceptibility testing (AST) relies on the bacterial culturing method, which often is cumbersome, time-consuming, and delays the therapeutic protocol. This chapter presents a brief overview of the bacterial structure, molecular mechanism of resistance in microbes, and various conventional and advanced molecular methods for testing antimicrobial susceptibility.

Keywords Microbial genomics · Antimicrobial susceptibility testing · Antimicrobial resistance · Antibiotics

D. Kalambhe (✉) · L. K.M. · G. Basak
Centre for One Health, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

S. Singh
Department of Veterinary Microbiology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

A. Jadhao
Department of Veterinary Pathology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_4

4.1 Introduction

With the discovery of microorganisms revealing the causes of many infectious diseases to the discovery of antimicrobial agents/antibiotics, which culminate such reasons, the journey may look fit and solution to the question. Antimicrobial resistance came into the picture within years of the discovery of the drugs leading to a great nuisance in the therapeutic world. In fact, with the passage of time, this resistance issue is instead increasing manyfold in terms of both frequency and variations. Therefore, proper point-specific detection of the cause becomes mandatory to culminate the same safeguarding the animal as well as public health. For specific targets, besides the phenotypic characteristics, genotypic properties are rather more prominent and promising too. Because the use of genome sequencing, epidemiological surveillance, and detection of resistance markers could be done at the same time, sequencing done in combination with bioinformatics can lead to the generation of systems that can replace classical methods. Antibiotic resistance is a major medical problem and rapid and precise detection of antibiotic resistance is of increasing importance. In a world where the spread of antibiotic resistance, both locally and internationally, appears to accelerate, there is an urgent need for more rapid diagnostic approaches. Phenotypic methods fulfill some of these needs, but speed is usually a limiting factor. Molecular testing should compensate for the time required for phenotypic testing. However, gene presence or mutation does not always correlate with phenotypic resistance. In addition, under antibiotic stress, bacteria will evolve, sometimes strain-dependent, beyond the 'usual suspect' genes which may confuse causal with random mutations. This indicates that research into new mechanisms of resistance to feed into the databases needs to be interpreted very carefully. Microbial genome sequencing might be a key player in the field of Antibiotic Sensitivity Testing (AST). However, there are still problematic subjects to be addressed, including the advancement in user-friendly data analysis, reduction of time and sequencing expenditures, and the continuous expansion of the databases with new resistance mechanisms and factors. In addition, clinical evaluation studies assessing the health effect of Whole-Genome Sequencing-based antibiotic therapy need to be executed on realistic patients. In fact, a genome sequence offers a comprehensive review of more than a few factors defining resistance.

4.2 Bacterial Structure

Bacterial cell structure is basically consisting of appendages (flagella, pili, or fimbriae), surface layers (capsule, cell wall, cell membrane), cytoplasm (nuclear material, ribosome, mesosome, inclusion body), and special structures (endospore). Glycocalyx is the outer viscous covering of fibers extending from the bacterial cell surface. Almost all bacteria secrete some. It consists of capsule and slime layer. Capsule is the extensive and tightly bound structure which is the accumulation of

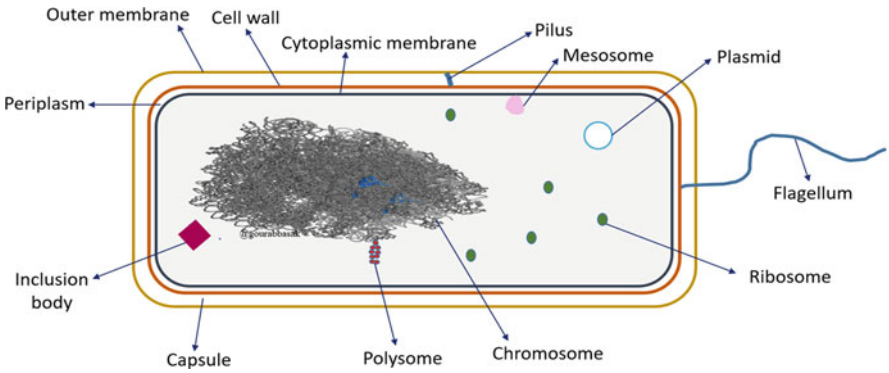

Fig. 4.1 Cell structure of a bacterium

gelatinous material adhered to cell wall, whereas slime layer is the unrecognized and more loosely attached part. Usually, glycocalyx is viscous polysaccharide but may also be polypeptide. Like in case of *B. anthracis,* it is poly-D-glutamic acid, peptidoglycan and muramic acid in *E. coli*, and hyaluronic acid in *Streptococcus pyogenes*. Mostly glycocalyx is comprised of water with <2% solids. Actual production depends upon environmental conditions, i.e., temperature, oxygen, ions, etc., like *B. anthracis* develops a capsule when growing in the body. Production may enhance when cells grow in sugar-rich medium. This functions as virulence determination by resisting phagocytic engulfment and saves engulfed bacteria from the attack of neutrophils by keeping away the lysosomal contents from the bacterial cell preventing killing, bacterial identification by K antigen, quelling reaction (using specific antiserum against capsular polysaccharide), mucoid colony characteristics on media, protection from desiccation, trapping of nutrients, biofilm formation, infection initiation by helping in adherence, colonization and resist flushing, vaccine development, and acts as an energy source in adverse conditions.

Flagellum, meaning 'whip,' is the filamentous protein structure attached to the cell surface. Flagella are distinct in form and function; they have evolved from archaeal and eukaryotic flagella, which are thinner than eukaryotic flagella. Also, there is a lack of typical 9 + 2 microtubule arrangement. Their length is several times of bacterial cells and their diameter is ~20 nm. Originating from the cell membrane, their number is 1–20 per cell, which is powered by proton motive force, i.e., chemiosmotic potential. Basically, flagella help in motility and swimming movement for most motile prokaryotes. Very few cocci, half of bacilli, and all spirals and/or curves are motile in nature. The ultrastructure of the flagellum is a complex structure of about two dozen proteins arranged into three parts, viz., a basal body, which is embedded in the cell wall and two axial filaments, hook and filament, which are joined at the hook-filament junction (Fig. 4.1). Besides, flagella can be peritrichous (*E. coli*, *Salmonella*), monotrichous (*Vibrio cholera*, *Pseudomonas aeruginosa*), lophotrichous (*Pseudomonas fluroscence*), amphitrichous (*Aquaspirillium*), and atrichous (*Shigella*).

Pilus, plural pili, from Latin meaning 'hair' and fimbria, plural fimbriae, from Latin meaning 'fringe', are the terms often used interchangeably. Originated from cell membrane, these are filamentous hair-like appendages, shorter, straighter, and thinner (4–8 nm) than flagella, peritrichous, and numbers are far more than flagella that may be up to 1000. The structure primarily composed of oligomeric protein, pilin which are arranged helically to form a cylinder, and can be dissociated into smaller units. Pilin basically belongs to a class of Lectin proteins that bind to the cell surface polysaccharide. The new protein subunits are inserted into the base of pilus. Moreover, pili are antigenic and fragile; these can constantly be replaced sometimes with different pili compositions leading to altered antigenicity. The specific host responses to old pili structure generally are not effective on the new recombination genes of pili code for variable (V) and constant (C) regions as in cases of Ig diversity. There are different types of pili present. Common pili is also known as adhesins that are numerous over the cell usually called fimbriae. It helps in adherence and enables bacteria to recognize specific receptor sites on host cell membrane and resists phagocytosis and pellicle formation in aerobic organisms. Virulence is one of the invasive mechanisms of bacteria. In fact, some pathogens cause diseases because of pili, whereas mutants which are non-fimbriate are non-virulent. Sex pili or conjugate pili are specialized kinds of pili forming attachments between male (doner) and female (recipient) cells, consisting of 1–4 per cell. These are essential for bacterial conjugation and the subsequent transfer of nucleic acids (DNA) from one cell to another. Many bacterial viruses infect only those bacteria which have F pilus.

Type IV pili facilitate many functions viz., locomotion, adherence, DNA uptake/competence, protein secretion, and biofilm production. Some also perform as nanowires carrying electric current. Basically, they are found throughout the gram-negative bacterial families, but are also widespread in gram-positives also. The external ends of pili adhere to solid substrate, where bacterium is attached or to other bacteria. On contraction of the pilus, the bacterium gets pull-forward showing typically a jerky movement which is also known as twitching motility. The type IV pilin structure is similar to flagellins of archaella (archaeal flagella). ComP pilin which aids in bacterial transformation is a type of type IV minor pilin. It can bind any DNA via an electropositive stripe exposed on the filament's surface, but has marked preference for *Neisseria meningitides* DNA Uptake Sequence (DUS) motifs.

Archaeal cell wall is composed of protein, polysaccharide, or peptidoglycan like molecules but never murein. The gram-positive's cell wall is 15–80 nm thick with several layers of peptidoglycan consisting 60–90% of cell wall. However, protein usually is not a constituent except M protein of Streptococci group. Importantly, teichoic acid which is anionic glycol-polymers is unique to gram-positives and constitutes 60% of the cell wall contributing normal functioning of bacterial cell. The teichoic acids of the wall are covalently linked to muramic acid residues in peptidoglycan by phosphodiester bonds which aid in cell division regulation, rigidity to cell wall, ion exchange, and antigenic determinants and help in antimicrobial resistance production. On the other hand, gram-negatives possess thin peptidoglycan layer constituting 10–20% of cell wall. They have additional membrane structure called outer membrane. The fluid-filled space between plasma membrane and outer

membrane is the periplasmic space that contains high concentration of degradative enzymes and transport proteins. The outer membrane is anchored to plasma membrane by Braun's lipoproteins and has unique asymmetric bilayer having outer lipopolysaccharide and inner phospholipids.

Cytoplasmic membrane or plasma membrane is thin structure of 7.5 nm inside the cell wall which is primarily made up of ~40% phospholipids and ~ 60% proteins for structural and enzymatic functions. It is the most dynamic structure. Phospholipids and proteins are not static and can move freely within membrane surface. The structure is well known as fluid mosaic model. The peripheral proteins reside on the outer and inner surface of the membrane which can be easily removed by mild treatment and function as enzymes, scaffolds for support. The integral proteins are usually transmembrane which can be removed only on disruption. However, some form porous channels. The plasma membrane contributes a definite cellular structure, molecular sequestration of molecules of life in the cytoplasm, and more importantly, provides partition from the outside environment. Besides, it mediates energy-generating functions, biosynthetic functions, chemotaxis and sensing functions, and transport. The bacterial transport processes consist of active and passive transports. Passive transport includes simple diffusion, osmosis, and facilitated diffusion.

4.3 Mechanism of Resistance

In order to smoothly understand how antimicrobial resistance occurs, it is required to know about how antimicrobials act on the target microbes first. Antimicrobials either function as cidal or static agents, i.e., either these can kill the target or inhibit pathogens. Generally, antimicrobials target cell wall, nucleic acid, membrane, protein synthesis, and nucleic acid synthesis. Hence, these are the following mechanisms of how they accomplish their work.

4.3.1 Cell Wall Synthesis Interference

Cell wall of bacteria guards the entire cell maintaining integrity and from any damages. This elastic macromolecule possesses peptidoglycan as the major component which is made up of long glycan chains of N-acetylglucosamine and N-acetylmuramic acid which is cross-linked by short peptides consisting of four amino acids. Transpeptidase and carboxypeptidase aid in this process. This is also known as Penicillin-Binding Proteins, being the binding target of various antibiotics, mostly β-lactams (penicillin, cephalosporin, carbapenem, monobactam) and glycopeptides (vancomycin, teicoplanin) (Tenover 2006; Johnson et al. 2013; Cho et al. 2014; Nikolaidis et al. 2014). The β-lactams resemble D-Ala-D-Ala dipeptide of nascent peptidoglycan and enable to bind to serine active site of Penicillin-Binding

Proteins which in turn hinder cross-linking of peptidoglycan in cell wall synthesis, whereas glycopeptides prevent link formation between peptidoglycan layer by acting on transpeptidase (Zapun et al. 2008; Cole and Riordan 2013).

4.3.2 Protein Synthesis Inhibition

Synthesis of protein is an essential and a must biological process which includes transcription and translation process involving four major steps of initiation, elongation, termination, and recycling. The structural difference of bacterial and eukaryotic ribosomes enables antibiotics to inhibit synthesis of proteins. Usually, antibiotics hinder the 30S or 50S subunit of 70S bacterial ribosome which slows or stops the growth of cells. Macrolides, aminoglycosides, and tetracyclines bind at 30S subunit and chloramphenicol binds at 50S subunit interrupting elongation step (Tenover 2006; Wilson 2014; Vazquez 1974).

4.3.3 Nucleic Acid Synthesis Inhibition

Topoisomerases is the principal enzyme of bacterial nucleic acid (DNA) synthesis (Hooper 2001). DNA 5 Gyrase and Topo IV which are type of IIA topoisomerases that helps in supercoiling which in its absence causes abnormality. Fluroquinolones inhibit DNA gyrase enzyme to play which is must for DNA replication, whereas inhibit topoisomerase IV function which is essential for decatenation (daughter-cell segregation) in gram-positive organisms (Blandeau 1999; Khodursky et al. 1995). Thus, quinolones alter DNA supercoiling binding to topoisomerase II or IV, leading to cell death by breaking of double-stranded DNA (Kohanski et al. 2010).

4.3.4 Metabolic Pathway Inhibition

Reduced folate cofactors are the key of manufacturing wide ranges of cellular components. The eukaryotic system consumes folate using active transport system and microbes utilized by de novo pathway. Thus, folate biosynthesis pathway becomes the target site of antibiotic action. In fact, para-aminobenzoic acid (PABA) is required by dihydropteroate synthase (DHPS) in the folate synthesis cycle and sulphonamides inhibit the usage of PABA (Woods 1940). Sulphonamides are competitive inhibitors as these possess structures similar to that of PABA (Roland et al. 1979).

4.3.5 *Bacterial Membrane Structure Interruption*

Plasma membrane and cell wall of bacterium constitute its cell envelope. Peptidoglycan which is present outside the cytoplasmic membrane serves as a permeability barrier. Moreover, bacteria are negatively charged as there lie peptidoglycan and lipopolysaccharides of outer membrane, whereas polymyxin antibiotics are positively charged. Thus, polymyxins easily bind with the outer membrane of the bacteria and alter the structure leading to extra permeable, osmotic imbalance, respiration prevention, and rapid water intake consequence in cell death (Fig. 4.2).

Antimicrobial resistance is the microbial ability to not getting inhibited by usually achievable antimicrobial concentration with normal prescribed dose-regiments and/or fall in minimum inhibitory concentration. Resistance to antimicrobial is basically of two types, viz., intrinsic resistance and acquired resistance.

Some specific bacterial genera or species provide natural resistance to certain antibiotics because of presence of some unique structural and/or functional characteristics. Such bacteria do not possess a target site for specific antibiotics making them ineffective leading to escaping of the bacteria. *Mycoplasma* spp. do not have cell wall which makes them naturally resistant to β-lactam antibiotics and glycopeptides (Abushaheen et al. 2020). Below is mentioned some examples of bacteria having intrinsic resistance against some antimicrobials they are for (Reygaert 2018).

- All Gram-positive organisms—aztreonam.
- All Gram-negative organisms—glycopeptides, lipopeptides.
- *Acinetobacter* spp.—ampicillin, glycopeptides.
- Bacteriodes (anaerobes)—aminoglycosides, quinolones, β-lactams (many).
- Enterococci—aminoglycosides, cephalosporins, lincosamides.
- *Escherichia coli*—macrolides.

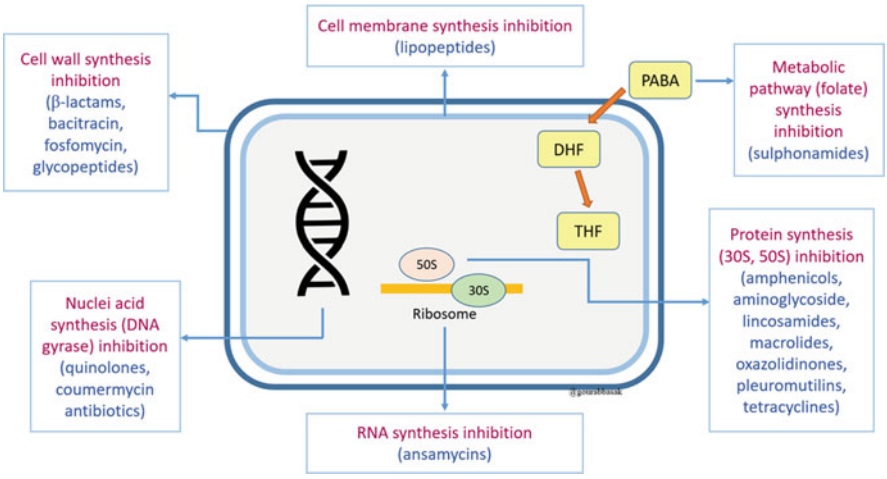

Fig. 4.2 Mechanism of action of different antibiotics on bacterial cell (Reygaert 2018)

- *Klebsiella* spp.—ampicillin.
- *Listeria monocytogenes*—cephalosporins.
- *Pseudomonas aeruginosa*—sulphonamides, ampicillin, chloramphenicol, tetracycline, first- and second-generation cephalosporins.
- *Serratia marcescens*—macrolides.
- *Stenotrophomonas maltophilia*—aminoglycosides, β-lactams, carbapenems, quinolones.

4.4 Acquired Resistance

Acquired resistance occurs when naturally susceptible bacteria can develop resistance against certain antibiotics by receiving genetic codes from other bacterial strains (Abushaheen et al. 2020). This can be achieved by the following three major mechanisms.

4.4.1 Reduced Intracellular Accumulation of Antimicrobial Agents

This again can be carried out in two ways, viz., limiting uptake of a drug and active efflux of a drug. The cell walls of gram-positive bacteria are permeable to most of the antibiotics but not the gram-negatives as the later form a permeability barrier because of the presence of LPS layer in their outer membrane that prevents antibiotics from entering the bacterial cell. Porins are the proteins forming channels embedded in the outer membrane of the gram-negative organisms. However, these porin channels are fairly nonspecific and can transport many antibiotics across the membrane. The modification of these porin channels can be achieved by mutation causing decrease in the number of porin channels and selectivity changes of the porin channels; thus, influencing the limited uptake of antimicrobial compounds. On the other hand, efflux is the movement of molecules out of the cell. Some antibiotics are efficiently removed by efflux using efflux pumps present in the membrane. Therefore, antibiotics' removal from the cell prevents the same from binding to its target, making the bacteria expressing efflux pumps resistant to antibiotics. Some pumps are specific, transporting only one class of antibiotic, whereas most of these are multidrug-resistant efflux pumps, mediating wide range of drug category. Some are expressed constitutively and some are induced or overexpressed. Most bacteria possess a variety of efflux pumps, but based on structure and energy sources, there are five main bacterial efflux pump families, viz., ATP-binding cassette (ABC) family, multidrug and toxic compound extrusion (MATE) family, small multidrug resistance (SMR) family, major facilitator superfamily (MFS), and resistance nodulation cell division (RND) family. Besides, porins and efflux pumps carry opposite reactions on

concentration of antibiotic in the cell. The number of both can be altered on the outer membrane of bacteria affecting the antibiotic concentration in the cell and ultimately the antibiotic susceptibility of bacteria (Poole et al. 2005).

4.4.2 *Modification of Enzymes or Inactivation of Antimicrobial Agents*

Both of these exercises are practiced by both gram-positive and negative bacteria. The enzymatic modification occurs by incorporation of acetyl, adenyl, or phosphate groups bacterial enzymes to specific antibiotic sites in order to modify the channel and subsequently antibiotic inactivation can be achieved by inability to bind with the target site. Aminoglycosides are associated with acetylation, adenylation and phosphorylation and macrolides with phosphorylation. In case of the enzymatic inactivation there occurs direct binding of the bacterial enzymes with antibiotics resulting in disintegration which is mediated by the hydrolytic cleavage action of antibiotic as like β-lactamases against penicillin and cephalosporins.

4.4.3 *Alterations at the Target Sites of Antimicrobial Agents*

Bacterial cell consists of numerous components which act as the targets for antimicrobial agents. Here, bacteria themselves change/modify these targets in order to enable resistance to these drugs. Topoisomerases type II in bacteria are essential enzymes of cellular processes including DNA replication. DNA gyrase and DNA topoisomerase IV which are hetero-tetrameric are also important players in this context. Type II topoisomerase comprises two copies of each, either GyrA and GyrB or ParC and ParE subunits, respectively. Though the enzymes possess homologous activity and even both DNA gyrae and topoisomerase IV can positively relax supercoiled DNA, only DNA gyrase is able to institute negative supercoils in relaxed DNA. Fluoroquinolones are the inhibitors of this type II topoisomerases of bacteria. They target both DNA gyrase and topoisomerase IV with varying degree in different bacteria inhibiting supercoiling within cell resulting in impaired DNA replication in low concentration and cell death at lethal concentrations. It depends on species of bacterial and specific fluroquinolones. Generally, in cases of gram-positive organisms, mostly topoisomerase IV is targeted and in case of gram-negative DNA gyrase is targeted. The most likely fluoroquinolone resistance lies with the mutation of genes encoding type topoisomerases, viz., GyrA, GyrB, ParC, and ParE, which alter the structure of target protein and fluoroquinolone-binding affinity of the enzyme leading to production of resistance (Fig. 4.3). Another prominent example is the mutation of 30S ribosomal subunit producing resistance to streptomycin in bacteria.

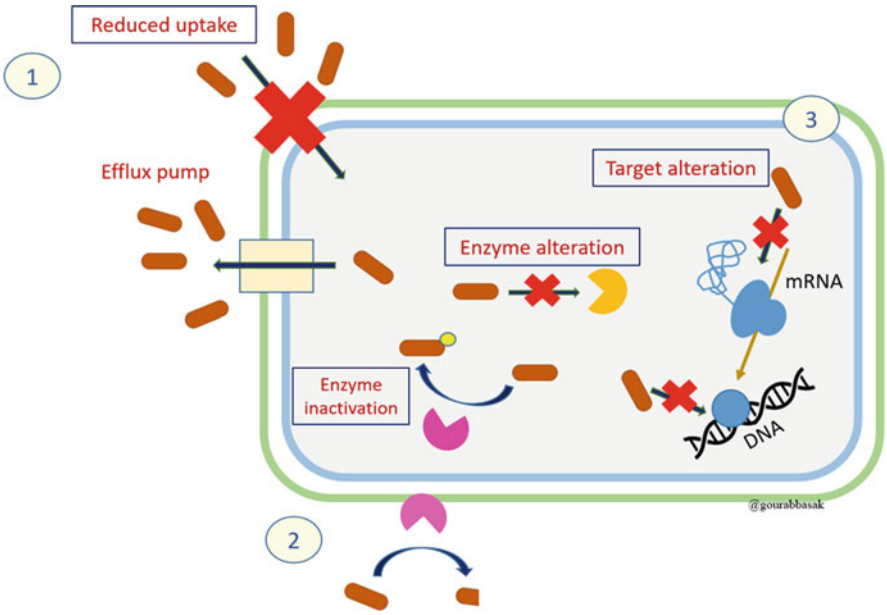

Fig. 4.3 Acquired resistance (Ali-Shtayeh and Abu Ghdeib 1999). Reduced intracellular accumulation of antimicrobial agents (Arikan 2007). Modification of enzymes or inactivation of antimicrobial agents (Baker et al. 1991). Alterations at the target sites of antimicrobial agents

Because of the differences in structures of gram-positive and gram-negative bacteria, different bacteria utilize different mechanisms. Gram-negative bacteria use all these three mechanisms, whereas positives use less commonly the limiting uptake of drug as they lack outer membrane in their structure and no capacity for drug efflux mechanisms.

Now, question arises here that how these resistance genes disseminate. Dissemination occurs among bacteria via three major processes, viz., bacterial transformation, bacterial transduction, and bacterial conjugation. Genetic transfer is the technique of efficient and stable introduction of foreign genes into genome of target cell. The transferred gene is known as transgenes and organism which receives a successful gene transfer is known as transgenic. Bacterial transformation is the transfer of relatively small segment of DNA from doner cells to the competent recipient cells. Transduction, on the other hand, is the transfer of DNA from one bacterium to another counterpart using bacteriophages. Transduction usually occurs through either lytic or lysogenic cycle. When the lysogen is induced in any way, e.g., U.V. light, phage genome from bacterial chromosome gets excised initiating the lytic cycle. This hence ceases in lysis of the cell and subsequent release of the phage particles. This cycle thus leads to generation of new phages which are released by the lysis of the host. In the lysogenic cycle, there is integration of phage chromosome into the bacterial chromosomes with the help of covalent bonds that remain in dormant stage for thousands of generations. Conjugation is another way of gene

transfer using sex pilus. For the accomplishment of conjugation, presence of special plasmid called F plasmid (contains 25 genes coding for sex pili production) is required. Bacteria possessing F plasmid are called F+ or male and those do not have are F- or female. Conjugation takes place on extension of sex pili of male and its attachment to the female counterpart resulting in two male cells. If F plasmid is found to be integrated in bacterial chromosome, term Hfr (high frequency of recombination) is used for the cell.

4.5 Antimicrobial Susceptibility Testing

Antimicrobial susceptibility testing (AST) uses various methods which are used to record the response of different microorganisms towards antimicrobial agents under controlled conditions. The ongoing problem of resistance has led to increase in frequency of empiric treatment with antibiotics like glycopeptides and broad-spectrum beta-lactams, for example, carbapenems and piperacillin–tazobactam. There has been increased appearance of multi-resistant microorganisms due to the excessive use of some reserved antimicrobial drugs. But these techniques suffer due to their prolonged operative durations as are dependent on growth measurement of microorganism either in the presence or absence of antimicrobials. The process follows overnight incubation generally, but sometimes the period is prolonged in cases of some anaerobes as they require additional time for growth under specific conditions. The methods usually prefer to afford qualitative results while using susceptible, intermediate, or resistant categories, but some of them also provide quantitative results which are expressed as minimum inhibitory concentration (MIC). The range of antimicrobial compound concentrations is generally performed in the form of two-fold dilutions. The specific MICs (breakpoints) are able to distinguish susceptible from resistant microorganisms which can vary according to the species and antibiotics. Upon comparison, the phenotypic AST is found to have two advantages ahead of the genotypic. The first advantage happens to be the ability to predict the pattern of drug resistance as well as the susceptibility and the second one being the ability to enlist the level of susceptibility of a particular pathogen towards any antimicrobial agent.

Any delay in appropriate antibiotic therapy can lead to prolonged hospitalization and increased patient mortality rate and is primarily related with increased costs. So, apart from these, alternative approaches need to be way out by the development of various rapid methods like high-speed molecular diagnostics. However, drawbacks like underestimating inducible mechanisms of resistance or insufficient time for detection of hetero-resistant subpopulations of these molecular methods need to be addressed and curbed accordingly for effective responses. Besides, downregulations like cost-effectiveness and gene function or regulation hindrance because of the presence or absence of a gene or a mutation act as speed-breaker in accurate prediction of antibiotic resistance. This may be because genes may not be expressed and mutations may be 100% silent.

For the purpose of improvement in sensitivity and speed of detection of bacterial resistance, novel techniques like several biosensors based on either chip calorimetry, electrical conductivity, millifluidic droplet analysis, or surface plasmon resonance have placed themselves in the list. But all these methods allow a single-sample analysis. Real-time polymerase chain reactions (PCR), microarrays, mass spectrometry, and flow cytometry have evolved recently due to their ability of high sensitivity, but at the same time they require special probes, expensive facilities, and technically qualified staff.

Some methodologies were also described which were categorized into both near-future alternative and long-term alternative. In the near-future alternative category, methods like cantilever technology, fluorescence-activated cell sorting, magnetic bead spin, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, microdroplets, and next-generation sequencing of all cellular DNA/RNA were described. The second group comprises of long-term alternative methods meant to be used in the future times and involve techniques like apoptosis markers, bacteriophage amplification, colorimetric detection of cell respiration, electronic noses, impedance measurements, infrared spectroscopy, liquid chromatography–electron spray ionization mass spectrometry, metabolomics including reactive oxygen species assessment, microsound measurements, nuclear magnetic resonance, Raman spectroscopy, and RNA sequencing.

In recent days, whole-genome sequencing (WGS) technique of cultured microorganisms or direct metagenomic sequencing of clinical materials has been advised as an alternate option to simulate phenotypic AST. It is not feasible at present to use this technology in all laboratories of microbiology as it requires facilities like high expenses, ease of use and automation, data interpretation and storage, curated resistance database development, etc.

Nucleic acid amplification of resistance genes or resistance-related genetic polymorphisms is mainly considered as a routine technology. This rapid technology is sensitive and specific, can be automated and rendered high-throughput, and can be performed directly on extracts of clinical materials. This may involve new technological platforms like loop-mediated isothermal amplification or the addition of new resistance markers. Identification of new resistance markers or development of molecular means of defining bacterial growth will speed up the process of AST by polymerase chain reaction (PCR).

Another approach is microbial genome sequencing for resistance estimation and direct care of the patient. For this alignment, annotation and the database used for making the calls need to be developed and quality controlled. Recently, comparing databases that have been developed within the framework of genomic AST is multifactorial. First, sequence-based AST is relatively new and existing databases have not been thoroughly evaluated as part of clinical trials or FDA submission studies. Second, the field of AST itself is evolving continuously with new resistance mechanisms being recognized on a near daily basis. This generates a highly dynamic and quite competitive research arena which is hard to evaluate in terms of verification and validation studies using well-defined strain collections.

ResFinder is one of the first examples of a web-based tool that could be used for tracking resistance genes in genomic databases which uses Basic Local Alignment Search Tool (BLAST) for detecting resistance genes and initially the database contained 1411 resistance genes recovered from 1862 GenBank genome files. The initial pipelines are being expanded and completed into integral systems for genomic sequence analysis for diagnosis and surveillance. An important additional genomic AST database is the Canadian Comprehensive Antibiotic Resistance Database (CARD).

Current databases are far from mature, and so, there are still many genetic mechanisms of resistance to be vetted. Further, there are likely to be markers of in vitro resistance that do not result in clinical expression of resistance. This type of discrepancy, falsely limiting the use of certain antimicrobial agents, would not be detrimental to the patient in terms of inappropriate therapy. With gram-negative bacilli, there are likely hundreds to thousands of resistance markers that are constantly evolving, so the databases will have to fulfill requirements like continuous updating, quality control, and validation. A number of other problems inherent to metagenomic prediction of resistance exist that include:

1. not all resistance mechanisms are expressed and genomic assessment of expression level is complicated,
2. some resistances are combinatorial, i.e., require multiple genes and/or mutations acting in concert,
3. metagenomic sequence analysis could identify resistance mechanisms produced by avirulent organisms.

However, using genome sequencing epidemiological surveillance and detection of resistance markers could be catalogued at the same time. Sequencing in combination with an adequate bioinformatics approach may generate systems that will in the end replace more classical approaches.

4.6 Antimicrobial Activity Evaluating Methods

There is availability of battery of methods for screening and quantifying the antimicrobial effects. The following includes different antibiotic sensitivity testing methods (Fig. 4.4).

4.6.1 Diffusion Methods

4.6.1.1 Agar Disk Diffusion Test

Since its development in 1940, this method is the largely and often used method in clinical microbiological laboratories (Heatley 1944). Though not all the fastidious

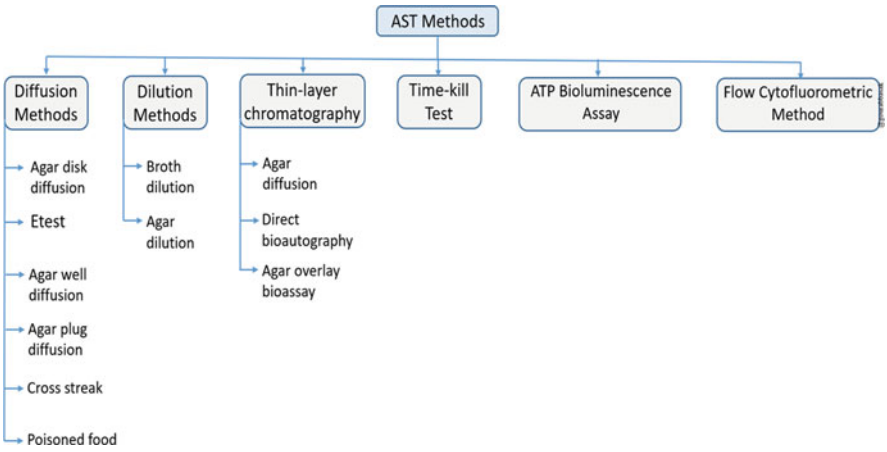

Fig. 4.4 Different methods of antimicrobial activity evaluation

organisms can be screened by this method, certain among them can be done now-a-days using specific culture media and employing various incubation conditions and imperious criteria for zone inhibitions, like in cases of *Haemophilus* spp., *Neisseria* spp., and Streptococci (CLSI 2012). It is initiated by inoculation of agar plates with standardized test organism's inoculum, followed by placement of 6 mm diameter filter paper disks upon agar surface containing the test compound at desired concentration and subjected for incubation. The antimicrobial agent gets diffused in the plate inhibiting the growth of the microorganisms. The diameter of the inhibition growth zones is then measured and the antibiogram inferred as susceptible, intermediate, and resistant on subsequent bacterial categorization which made it a typical tool based on microbial resistance phenotype (Jorgensen and Ferraro 2009). However, one of the major drawbacks is the non-appropriate determination of minimum inhibitory concentration (MIC) as it is not possible to determine how much amount of antimicrobial agent used get dissociated in the media (Nijs et al. 2003).

4.6.1.2 Antimicrobial Gradient Method/Etest

It is the combined principle of both dilution and diffusion method in order to have the MIC value which is based on concentration gradient of the antimicrobial agent tested in agar medium. Impregnated strip with increasing concentration gradient of antimicrobial agent is placed over the inoculated agar surface. The MIC is determined at the intersection of the strip and growth inhibition eclipse. Because of its simple methodology, it is commonly used for MIC determination of antibiotics, antifungals, and antimycobacterial (Hausdorfer et al. 1998). One advantage of it is that it enables to find out antimicrobial interaction between two drugs (White et al. 1996).

4.6.1.3 Agar Well Diffusion Test

Like disk diffusion method, for this method also microbial inoculum is spread over the agar plate surface and 20–100 μl of antimicrobial agent is introduced into a hole of diameter 6–8 mm which is punched aseptically on it. The plates are incubated accordingly based on the organism chosen. The antimicrobial agent diffuses and prevents the growth of the concerned microbe. This method is widely used in cases of plants or microbial extracts' antimicrobial activity determination (Magaldi et al. 2004).

4.6.1.4 Agar Plug Diffusion Method

Resembling to the disk diffusion method, this method involves making of agar culture medium of bacterium of interest by streaking. During incubation and their growth in the plate, they secrete some molecules which diffuse on the plate. After incubation, an agar-plot or cylinder is cut from the same and placed on another plate which is previously inoculated with the test organism where that substance will diffuse in the medium from that plot or cylinder. Hence, the antimicrobial property of the secreted molecule can be detected by the zone of inhibition appeared in the plate. This method is useful to highlight the antagonism between organisms (Jiménez-Esquilín and Roane 2005).

4.6.1.5 Cross Streak Method

This is another method for testing microbial antagonism by single streak of the bacterium of interest on the center of the plate and incubated. After incubation, the test organism is streaked in the center perpendicular with the previous streak. On incubation, the antimicrobial activity can be analyzed by zone of inhibition size (Lertcanawanichakul and Sawangnop 2008).

4.6.1.6 Poisoned Food Method

Antifungal agents are best evaluated by this method. These agents being incorporated in the molten agar at desired concentration and poured into plates on proper mixing. After overnight preincubation, inoculation at the center of the plate is carried out using mycelia disc of 2–5 mm diameter. On subsequent incubation for the desired fungal strain, their growth diameters in control and the sample plates are measured and antifungal activity is measured using the formula: Antifungal activity (%) = [(Dc−Ds)/Dc] × 100 where Dc is growth diameter in control plate and Ds is growth diameter in tested plate. Though the process of sporulation can also be compared with the control (Ali-Shtayeh and Abu Ghdeib 1999).

4.6.2 Dilution Methods

4.6.2.1 Broth Dilution Method

Broth dilution including both micro and macro dilutions is the basic AST method which involves preparation of two-fold dilutions of the antimicrobial agents in liquid medium. For microdilution, a minimum of 2 mL of broth dispensed in tubes is required whereas for microdilution, smaller volumes using 96 well microtitration plate is required; in which microbial inoculation is made after standardized microbial suspension diluted to 0.5 McFarland scale and subjected for incubation. MIC which is the lowest concentration of antimicrobial agents completely inhibiting the growth of microorganisms detected by naked eye, its endpoint for microdilution is aided by viewing devices. Besides, various calorimetric methods using dyes can also efficiently be used in MIC endpoint determination. Microdilution is superior than microdilution because the latter is tedious to work and manually operated imposing risk of errors during solution preparations for each test with comparatively requirement of higher amounts of reagents and space (Jorgensen and Ferraro 2009). In place reproducibility of the test and lesser reagent usage make the former more advantageous (Reis et al. 2004). For preventing influence of MIC values (from inoculum size, media type, incubation time, inoculum preparation etc.), CLSI has standardized broth dilution for aerobic bacteria, yeast and filamentous fungi growth. EUCAST has also prescribed broth dilution method which is almost similar to that of CLSI with little modifications (Arikan 2007).

4.6.2.2 Agar Dilution Method

This technique is suitable for both antibacterial and antifungal susceptibility analysis as it involves the incorporation of desired antimicrobial agent concentrations in molten agar on serial two-fold dilution and subsequent bacterial inoculation on the media. After incubation, the MIC endpoint is determined as the agents completely inhibiting the microbial growths. Moreover, this method gives a better correlation with Etest employed against gram-positive and negative microorganisms (Baker et al. 1991).

4.6.3 Thin Layer Chromatography

4.6.3.1 Agar Diffusion Method

Agar diffusion or agar contact method involves the diffusion transfer of antimicrobial agent to agar plate already inoculated with the test microorganism from the chromatogram (TLC or P.C.) and after few minutes to hours of diffusion, plates are allowed for incubation on removal of the chromatogram. After incubation, inhibition

zones appear in the plate where the antimicrobial agents lay with the agar. This is the least applied method as compared to others (Marston 2011).

4.6.3.2 Direct Bioautography

Among the three thin-layer chromatography (TLC) methods, direct bioautography is the most applied one. To accomplish the procedure, TLC plates are required which are dipped or sprayed by microbial suspension and incubated under humidity at 25 °C for 2 days (Dewanjee et al. 2015). Thereafter tetrazolium and/or p-Iodonitrotetrazolium violet salts are used for visualization, among which later is the most suitable detector (Choma and Grzelak 2011). These salts are sprayed over the bioautogram and incubated at either 25 °C for a day or 37 °C for 3–4 h (Silva et al. 2005). This method is utilized for both fungi and bacteria. In fact, gives the promising results and easier for the antifungal compounds and spore producing fungi (Suleiman et al. 2010).

4.6.3.3 Agar Overlay Bioassay

The mixture of agar diffusion method and direct bioautography is the agar overlay or immersion bioautography method which can be applied to various microorganisms including Candida and moulds as like direct bioautography (Balouiri et al. 2015). Here, the TLC plates are covered with molten agar and incubated at low temperatures for a few hours for better results. On incubation staining is carried out with tetrazolium dye under suitable condition based on the test organism. The best part is it is less sensitive to contamination and gives prominent zones of inhibitions (Marston 2011).

4.6.4 Time-Kill Curve/Test

This method defines either time-dependent or concentration-dependent antimicrobial effects highlighting the interaction between antimicrobial agent and the respective microbe involved. Thus, is a suitable determining method for bactericidal and fungicidal effects which is performed in broth cultures having bacterial suspension of 5×10^5 CFU/mL. It uses three tubes, the first one contains a concentration of tested molecule to be 0.25 × MIC, the second one possesses concentration of 1 × MIC and the third one is taken as control. The incubation is carried out at different time intervals as 0, 4, 6, 8, 10, 12 and 24 h (Pfaller et al. 2004). Thereafter, dead cells percentage is calculated relatively to control. A lethal percentage of 90% for 6 h or 99.9% for 24 h (equivalent to the former) is attained as the bactericidal effect (Konaté et al. 2012).

4.6.5 ATP Bioluminescence Assay

As the name indicates, the principle of the assay is to measure the ATP produced by bacteria or fungi and its subsequent light production. Thus, its quantification helps to estimate the microbial load in a particular sample. D-luciferin in the presence of ATP converts to oxyluciferin which emits light. Now the quantity of light emitted is measured using luminometer which is expressed as relative light unit (RLU); further, it is converted to RLU/mole of ATP giving a linear relation in between the viability of cell and the luminescence measured. The bioluminescence assay provides many advantages, including its usage in in vivo and in situ antimicrobial testing (Vojtek et al. 2014).

4.6.6 Flow Cytofluorometric Method

This method is depended upon the usage of appropriate dye staining in order to detect the damaged cells (Ramani et al. 1997). A fluorescent and intercalating agent, propidium iodide (P.I.), is commonly used as a DNA strain (Ramani and Chaturvedi 2000). Apart from the lysed cells, dead, viable and injured cells are also appropriately distinguished. Similarly, antimicrobial resistance detection and estimation of tested molecular impact on tested microbial viability and cell damage can be carried out. Though it provides rapid results within 2–6 h its use in AST is unlikely, and the facility of flow cytometry is limited in many laboratories (Ramani and Chaturvedi 2000).

4.7 Automated and Semiautomated Devices

For rapid and reliable bacterial identification and antibiotic sensitivity testing, these devices fulfill their goal prominently. These are advantageous because of cost-effectiveness, decrease turnaround times, convenient antibiotic selection, monitoring of antibiotic resistance pattern, assisting in disease diagnosis. This includes various techniques, viz., VITEK systems, MicroScan WalkAway system, Trek diagnostic systems, and B.D. phoenix system. VITEK device is used for identification as well as AST of both gram-negative and positive organisms by its automated photometric mechanism. It basically works on fluorescence, turbidity, and colorimetric signals (Ligozzi et al. 2002). VITEK 2 system is the second generation of the previous one. AutoSCAN-3, a semiautomated instrument, was the first generation of MicroScan WalkAway system which requires microdilution trays containing frozen substrates. The next-generation model, autoSCAN-4, uses dry panels which are not refrigerated. The current MicroScan WalkAway uses fluorogenic substrates and pH indicators for the detection of bacterial enzymatic activity (Rodriguez et al. 1999). Likewise, Trek

diagnostic system has two configurations, viz., Sensititre AutoReader and Sensititre ARIS 2X. The former is a fully automated fluorometer reader which is controlled by microprocessor and the latter seems to be fully automated overnight benchtop incubating and reading system. This instrument uses bar codes for identifying panels which are available for both gram organisms (O'Hara 2005). Again, the B.D. Phoenix system functions by adjusting the turbidity of bacterial suspension to McFarland standard of 0.5, preparing a dilution and addition of AST indicator to the inoculum for susceptibility testing. Thereafter, the panels used are monitored by colorimetric and fluorescence readings for bacterial identification and redox and turbidimetric detection for AST; the MIC results are available within 6–16 h (Jorgensen and Ferraro 2009).

4.7.1 Isothermal Microcalorimetry

This works under the principle of cumulative heat production, which increase or decrease, keeping in pace with bacterial growth curve and correspondingly with lag, log, and stationary phases (von Ah et al. 2009).

4.7.2 MALDI-TOF-MS

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry works on the principle of determination of mass/charge (m/z) ratio and strain differentiation using their ribosomal proteins and not dependent on metabolic activities (Maxson et al. 2017).

4.7.3 Microfluidic Devices and Microdroplets

These techniques are based on microchannels (10–100 μm); extremely small volumes of reagent and analyte in quantities of nanoliters or picoliters (Zhu et al. 2018; Hayat and El Abed 2018) can flow through the entire channels on one chip. These detection methods depend upon the type of device used and can vary widely and are generally electrochemical, magnetic, or optical/microcalorimetric (Schumacher et al. 2018). Microfluidic AST methods have comparatively a large potential for the purpose of application in clinical laboratories.

4.7.4 Polymerase Chain Reaction-Mediated Methods

The PCR-based genomic technologies are sensitive and specific if conditions are fulfilled. Also, they can be automated, render high-throughput, and enable to perform directly on extracts of clinical materials rapidly. Identification of new density nucleic acid sequencing provides better diagnostics. However, PCR testing only finds what is being specifically aimed for and a broader, sequencing-based approach has advantages. Hence, genomic AST is currently competing with targeted molecular approaches. However, both conventional and real-time are helping in detecting the presence and absence of genetic determinants of cause of resistance to different antimicrobial agents (Fluit et al. 2001).

4.7.5 Bioinformatic Pipelines for Genomic AST

For this approach, alignment, annotation, and databases used for making the calls need to be developed and quality-controlled. Comparing databases that have been developed recently within the framework of genomic AST is multifactorial. First, sequence-based AST is relatively new and existing databases have not been thoroughly evaluated as part of clinical trials or FDA-related studies. Second, the field of AST itself is evolving continuously with new resistance mechanisms being recognized on a near daily basis. This generates a highly dynamic and quite competitive research arena which is hard to evaluate in terms of verification and validation studies using well-defined strain collections. Whether such databases should be available for open public access or more restricted proprietary use is subject to debate including assigning responsibilities for database maintenance and quality assurance and control.

4.8 Other Approaches to Address AMR

1. Antibiotic Discovery.
   (a) New methods of natural antibiotic identification, viz., iChip technology.
   (b) Novel sources of natural products like marine microbes living in extreme conditions.
   (c) Bacteriocins like antimicrobial peptides.
2. Harnessing immunity like vaccines against major health care-associated infections, monoclonal antibody, innate immune modulators enhancing beneficial or suppressing deleterious immune responses like in cases of molecules modulating LPS interactions and toll-like-receptor 4.

3. Microbial community tweaking with bacteriophages, live biotherapies, and fecal microbiota transplant.
4. Antivirulence strategies like usage of secretion system inhibitors, antitoxin antibodies (toxins), application of quorum sensing, and biofilm formation.
5. Rapid detection including biomarkers like serum procalcitonin indicating bacterial infection, culture-independent methods like magnetic resonance technology, transcriptional profiling, and specific point of resistance detection.

References

Abushaheen MA, Fatani AJ, Alosaimi M, Mansy W, George M, Acharya S, Rathod S, Divakar DD, Jhugroo C, Vellappally S, Khan AA (2020) Antimicrobial resistance, mechanisms and its clinical significance. Dis Mon 66(6):100971. https://doi.org/10.1016/j.disamonth.2020.100971

Ali-Shtayeh MS, Abu Ghdeib SI (1999) Antifungal activity of plant extracts against dermatophytes. Mycoses 42:665–672

Arikan S (2007) Current status of antifungal susceptibility testing methods. Med Mycol 45:569–587

Baker CN, Stocker SA, Culver DH et al (1991) Comparison of the ETest to agar dilution, broth microdilution, and agar diffusion susceptibility testing techniques by using a special challenge set of bacteria. J Clin Microbiol 29:533–538

Balouiri M, Bouhdid S, Harki E et al (2015) Antifungal activity of Bacillus spp. isolated from Calotropis procera AIT. Rhizosphere against Candida albicans. Asian J Pharm Clin Res **8**:213–217

Blandeau JM (1999) Expanded activity and utility of the new fluoroquinolones: a review. Clin Ther 21(1):3–40

Cho H, Uehara T, Bernhardt TG (2014) Beta-lactam antibiotics induce a lethal malfunctioning of the bacterial cell wall synthesis machinery. Cell 159(6):1300–1311

Choma IM, Grzelak EM (2011) Bioautography detection in thin-layer chromatography. J Chromatogr A 1218:2684–2691

CLSI (2012) Performance standards for antimicrobial disk susceptibility tests, approved standard, 7th edn, CLSI document M02-A11. Clinical and laboratory standards institute, Wayne

Cole TS, Riordan A (2013) Vancomycin dosing in children: what is the question? Arch Dis Child 98(12):994–997

Dewanjee S, Gangopadhyay M, Bhattabharya N et al (2015) Bioautography and its scope in the field of natural product chemistry. J Pharm Anal 5:75–84

Fluit AC, Visser MR, Schmitz FJ (2001) Molecular detection of antimicrobial resistance. Clin Microbiol Rev 14:836–871. https://doi.org/10.1128/CMR.14.4.836-871.2001

Hausdorfer J, Sompek E, Allerberger F et al (1998) E-test for susceptibility testing of mycobacterium tuberculosis. Int J Tuberc Lung Dis 2:751–755

Hayat Z, El Abed AI (2018) High-throughput optofluidic acquisition of microdroplets in microfluidic systems. Micromachines 9:183. https://doi.org/10.3390/mi9040183

Heatley NG (1944) A method for the assay of penicillin. Biochem J 38:61–65

Hooper DC (2001) Emerging mechanisms of fluoroquinolone resistance. Emerg Infect Dis 7(2):337

Jiménez-Esquilín AE, Roane TM (2005) Antifungal activities of actinomycete strains associated with high-altitude sagebrush rhizosphere. J Ind Microbiol Biotechnol 32:378–381

Johnson JW, Fisher JF, Mobashery S (2013) Bacterial cell-wall recycling. Ann N Y Acad Sci 1277(1):54–75

Jorgensen JH, Ferraro MJ (2009) Antimicrobial susceptibility testing: a review of general principles and contemporary practices. Clin Infect Dis 49:1749–1755

Khodursky AB, Zechiedrich EL, Cozzarelli NR (1995) Topoisomerase IV is a target of quinolones in Escherichia coli. Proc Natl Acad Sci U S A 92(25):11801–11805

Kohanski MA, Dwyer DJ, Collins JJ (2010) How antibiotics kill bacteria: from targets to networks. Nat Rev Microbiol 8(6):423

Konaté K, Mavoungou JF, Lepengué AN et al (2012) Antibacterial activity against β-lactamase producing methicillin and ampicillin-resistant Staphylococcus aureus: fractional inhibitory concentration index (FICI) determination. Ann Clin Microbiol Antimicrob 11:18

Lertcanawanichakul M, Sawangnop S (2008) A comparison of two methods used for measuring the antagonistic activity of bacillus species. Walailak J Sci Technol 5:161–171

Ligozzi M, Bernini C, Bonora MG, de Fatima M, Zuliani J, Fontana R (2002) Evaluation of the VITEK 2 system for identification and antimicrobial susceptibility testing of medically relevant gram-positive cocci. J Clin Microbiol 40:1681–1686. https://doi.org/10.1128/JCM.40.5.1681-1686.2002

Magaldi S, Mata-Essayag S, HartungdeCapriles C et al (2004) Well diffusion for antifungal susceptibility testing. Int J Infect Dis 8:39–45

Marston A (2011) Thin-layer chromatography with biological detection in phytochemistry. J Chromatogr A 1218:2676–2683

Maxson T, Taylor-Howell CL, Minogue TD (2017) Semi-quantitative MALDI-TOF for antimicrobial susceptibility testing in Staphylococcus aureus. PLoS One 12:e0183899. https://doi.org/10.1371/journal.pone.0183899

Nijs A, Cartuyvels R, Mewis A et al (2003) Comparison and evaluation of Osiris and Sirscan 2000 antimicrobial susceptibility systems in the clinical microbiology laboratory. J Clin Microbiol 41:3627–3630

Nikolaidis I, Favini-Stabile S, Dessen A (2014) Resistance to antibiotics targeted to the bacterial cell wall. Protein Sci 23(3):243–259

O'Hara CM (2005) Manual and automated instrumentation for identification of Enterobacteriaceae and other aerobic gram-negative bacilli. Clin Microbiol Rev 18:147–162. https://doi.org/10.1128/CMR.18.1.147-162.2005

Pfaller MA, Sheehan DJ, Rex JH (2004) Determination of fungicidal activities against yeasts and molds: lessons learned from bactericidal testing and the need for standardization. Clin Microbiol Rev 17:268–280

Poole K (2005) Efflux-mediated antimicrobial resistance. J Antimicrob Chemother 56(1):20–51. https://doi.org/10.1093/jac/dki171

Ramani R, Chaturvedi V (2000) Flow cytometry antifungal susceptibility testing of pathogenic yeasts other than Candida albicans and comparison with the NCCLS broth microdilution test. Antimicrob Agents Chemother 44:2752–2758

Ramani R, Ramani A, Wong SJ (1997) Rapid flow cytometric susceptibility testing of Candida albicans. J Clin Microbiol 35:2320–2324

Reis RS, Neves I, Lourenço SLS et al (2004) Comparison of flow cytometric and alamar blue tests with the proportional method for testing susceptibility of mycobacterium tuberculosis to rifampin and isoniazid. J Clin Microbiol 42:2247–2248

Reygaert WC (2018) An overview of the antimicrobial resistance mechanisms of bacteria. AIMS Microbiol 4(3):482–501. https://doi.org/10.3934/microbiol.2018.3.482

Rodriguez LA, Vivas J, Gallardo CS, Acosta F, Barbeyto L, Real F (1999) Identification of hafnia alvei with the MicroScan WalkAway system. J Clin Microbiol 37:4186–4188

Roland S, Ferone R, Harvey RJ, Styles V, Morrison RW (1979) The characteristics and significance of sulfonamides as substrates for Escherichia coli dihydropteroate synthase. J Biol Chem 254(20):10337–10345

Schumacher A, Vranken T, Malhotra A, Arts JJC, Habibovic P (2018) In vitro antimicrobial susceptibility testing methods: agar dilution to 3D tissue-engineered models. Eur J Clin Microbiol Infect Dis 37:187–208. https://doi.org/10.1007/s10096-017-3089-2

Silva MTG, Simas SM, Batista TG et al (2005) Studies on antimicrobial activity, in vitro, of Physalis angulata L. (Solanaceae) fraction and Physalin B bringing out the importance of assay determination. Memorias do Instituto Oswaldo Cruz 100:779–782

Suleiman M, McGaw L, Naidoo V et al (2010) Detection of antimicrobial compounds by bioautography of different extracts of leaves of selected South African tree species. Afr J Tradit Complemen Altern Med 7:64–78

Tenover FC (2006) Mechanisms of antimicrobial resistance in bacteria. Am J Med 34(5):S3–S10

Vazquez D (1974) Inhibitors of protein synthesis. FEBS Lett 40(S1):S48–S62

Vojtek L, Dobes P, Buyukguzel E et al (2014) Bioluminescent assay for evaluating antimicrobial activity in insect haemolymph. Eur J Entomol 111:335–340

von Ah U, Wirz D, Daniels A (2009) Isothermal micro calorimetry - a new method for MIC determinations: results for 12 antibiotics and reference strains of E. coli and S. aureus. BMC Microbiol 9:106. https://doi.org/10.1186/1471-2180-9-106

White RL, Burgess DS, Manduru M et al (1996) Comparison of three different in vitro methods of detecting synergy: time-kill, checker board, and Etest. Antimicrob Agents Chemother 40:1914–1918

Wilson DN (2014) Ribosome-targeting antibiotics and mechanisms of bacterial resistance. Nat Rev Microbiol 12(1):35

Woods DD (1940) The relation of p-aminobenzoic acid to the mechanism of the action of sulphanilamide. Br J Exp Pathol 21(2):74

Zapun A, Contreras-Martel C, Vernet T (2008) Penicillin-binding proteins and β-lactam resistance. FEMS Microbiol Rev 32(2):361–385

Zhu XD, Chu J, Wang YH (2018) Advances in microfluidics applied to single cell operation. Biotechnol J 13:1700416. https://doi.org/10.1002/biot.201700416

Chapter 5
Genome-Wide Copy Number Variation and Structural Variation: A Novel Tool for Improved Livestock Genomic Selection

Wenli Li

Abstract Copy number variations (CNVs) are a major form of genomic variation in mammalian genomes. Due to their much larger size than the single nucleotide polymorphisms (SNPs), they affect a significant portion of the genome via deletion, duplication, and translocation. SNPs have been the variant of choice in genomic prediction in livestock species. With the advancement of sequencing technologies and analytical algorithms, significant number of CNVs have been associated with desirable production traits and disease phenotype, making this form of genomic variants an attractive candidate for improved livestock genomic selection. In this review, the current CNVs associated with livestock production traits and diseases and the obstacles of using CNV as a genomic selection were discussed. Finally, this review discusses the future of CNV identification and its application to improve livestock genomic selection.

Keywords Copy number variation · Livestock species · Genomic selection

5.1 Introduction

As a major form of genomic variation in mammalian genomes, CNVs are generated through duplication or deletion events that affect one or more loci. They range between 50 bp and several Mbps in length (Girirajan et al. 2011; Mills et al. 2011; Iskow et al. 2012; Zarrei et al. 2015). Genomic regions that are affected by CNVs are far greater than those with SNPs (Lupski 2007). About 15% of the human genome is affected by CNVs (Stankiewicz and Lupski 2010), and an average of 12 CNVs exist per individual compared with the reference genome. A substantial portion of CNVs is likely to have functional consequences through changing expression levels of genes in or near the affected regions due to positional effects, gene dosage alteration, and disruption or fusion of genes.

W. Li (✉)
USDA-ARS, Cell Wall Biology and Utilization Research, Madison, WI, USA
e-mail: wenli.li@usda.gov

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_5

However, the full impact of this "new" form of genome variation on genome function and dysfunction is still unclear. In the human and rodent genomes, CNVs have been well studied and have been associated with some phenotypic traits and diseases, adaptation, domestication, and evolution in humans and animals (Girirajan et al. 2011, 2013; Cook Jr. and Scherer 2008; Almal and Padh 2012; Redon et al. 2006; Innan and Kondrashov 2010; Shwan et al. 2017; Bayes et al. 2003; Cumer et al. 2021). In comparison, CNVs and their functional roles are less understood in livestock species. This form of genomic variants may be a novel genomic selection tool used for production trait improvement in livestock species.

5.2 The Creation of CNVs and Their Distribution in the Genome

CNVs arise through various mechanisms, and they fall into two major types: recurrent and nonrecurrent. Nonallelic homologous recombination (NAHR) is the mechanism that generates recurrent CNVs. Repeated sequences, including low-copy repeats (e.g., segmental duplications) and high-copy repeats (e.g., SINEs, LINEs, and endogenous retroviruses), are enriched in the vicinity of breakpoints and represent an important factor for CNV instability (Chen et al. 2014). These sequence motifs play a key role in NAHR, creating recurrent CNVs. Recurrent CNVs share a common size and have clustering of common breakpoints. Nonrecurrent CNVs can be formed by nonhomologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ), and replication errors brought about by fork stalling or template switching (Hastings et al. 2009). Nonrecurrent CNVs may share the smallest region of overlap. The mutation rate of CNVs is considered two to four orders of magnitude greater than point mutations (Hastings et al. 2009). CNV distribution hotspots have been reported in the human genome (Fu et al. 2010). Since CNVs cover more total bases, they are more likely to affect gene structure, regulation, and gene expression (Geistlinger et al. 2018). The following three categories of genes are enriched for structural variants: (1) genes involved in immunity and signaling pathways (Memon et al. 2021; Moasser 2007; Frenkel et al. 2020); (2) genes encoding proteins involved in interaction with the environment (e.g., immune response, drug metabolism, and perception of smell) (Schaeffeler et al. 2003; Young et al. 2008); and (3) retrovirus- and transposition-related protein-coding genes (Belshaw et al. 2004; Viginier et al. 2012).

In contrast, CNV occurrence is significantly reduced in genes that are widely expressed in different tissues (Henrichsen et al. 2009a). Dosage-sensitive genes are underrepresented in regions affected by CNVs (Schuster-Bockler et al. 2010). Functionally constrained genes appear to have fewer CNVs, likely due to the effect of purifying selection in removing structural variants in these regions (Dopman and Hartl 2007). Similarly, noncoding intronic regions had more frequent CNVs than the protein-coding regions (Rigau et al. 2019).

5.3 CNVs in Livestock Genetics

A key focus of livestock genetics is the identification and application of genomic variants with causal impacts on the individual phenotypes (Clop et al. 2012). Single nucleotide polymorphisms (SNPs) were considered an important cause of genetic variation by the livestock research community. Many studies have demonstrated the utilization and efficiency of SNPs for the trait improvement in beef and dairy cattle (Hayes et al. 2009; Li et al. 2018; VanRaden et al. 2009).

CNVs span considerately large area of the genomes and have a large-scale impact on the expression of the transcriptome (Henrichsen et al. 2009b). Increasing number of studies have reported the role of CNVs in generating essential variation in livestock population and phenotypic traits (Gorla et al. 2017; de Almeida Santana et al. 2016; Bickhart and Liu 2014; Durkin et al. 2012; Brenig et al. 2013; Upadhyay et al. 2017; Butty et al. 2020). Due to their large size, CNVs are considered important for genetic variation formation and adaptation (Liu et al. 2009; Hou et al. 2011; Zhang et al. 2016). Thus, the association of CNVs with phenotypic traits can be leveraged for the selection of desirable traits in livestock.

The first steps in applying CNVs as potential genetic markers are the development of CNV maps and their causal association with pertinent traits in livestock species. Several methods have been used for the generation of CNV maps in various livestock species. Using a 600 K SNP array, Wang and coauthors reported the genome-wide CNVs and their association with body weight in sheep (Wang et al. 2020). In this study, a total of 919 CNVs were detected in the sheep genome, spanning a total length of 48.17 Mbp and accounting for 1.96% of the entire genome. Functional analysis indicated that these CNVs were enriched in the olfactory transduction pathway. For the genes affected by the identified CNVs, many of them were linked to body weight, including *FOXF2*, *MAPK12*, *MAP 3 K11*, *STRBP*, and *C14orf132*. Using a BovineHD SNP array, Zhang and coauthors (Zhang et al. 2019) performed genome-wide CNV identification from 24 Chinese indigenous cattle breeds with 37 yaks as the outgroup. A total of 5818 autosomal CNVs, covering 14.34% of the bovine genome, was identified.

5.4 The Association Between CNVs and Production Traits in Livestock Species

CNVs are widespread in cattle breed (Upadhyay et al. 2017; Hou et al. 2011; Liu et al. 2010; Choi et al. 2013; Sasaki et al. 2016; Pierce et al. 2018; Rafter et al. 2018). Many studies have revealed the significant association between CNVs and important traits in cattle production. Using a cohort of 528 Holstein cows, Zhou et al. reported that a single CNV overlapping with an olfactory receptor gene (OR2A2) is associated with feed efficiency and intake-related phenotypes (Zhou et al. 2018). Other genes impacted by copy number variation have been linked to growth traits in cattle,

including LEPR (Shi et al. 2016; Xu et al. 2017), *GBP2* (Zhang et al. 2018), *MYH3* (Xu et al. 2014a) and *MAPK10* (Liu et al. 2016), *HPGDS* (Huang et al. 2021), and *CHRM3* (Goshu et al. 2020).

Using a whole genome sequencing approach, Gao and coauthors (Gao et al. 2017) compared the genomes of two groups of dairy cattle with high- or low-milk protein and milk fat percentage. Among the CNVs that differed between these two groups, 235 functional genes were identified. Interesting, these genes were significantly enriched in the pathways related to lipid and protein metabolism. Additionally, 95 of the CNVs overlapped with 75 known QTLs associated with the protein and fat traits in dairy cattle. Other production traits also showed significant association with CNVs, including milk composition (Gao et al. 2017), milk production (Xu et al. 2014b), meat tenderness (Silva et al. 2016), and reproduction (Liu et al. 2019).

Production traits-associated CNVs have been investigated in other livestock species, including goats, sheep, chicken, and swine. In dairy goats, Kang and coauthors reported the CNVs overlapping with *ADAMTS20* and *PAPPA2* could potentially have functional impacts in goat dairy traits (Kang et al. 2020). In sheep, the duplication of the *ASIP* gene was linked to the coat color (Norris and Whan 2008), and the copy number variations of *TOP2B* (Toremurat et al. 2022) and *KMT2D* (Cheng et al. 2020) were associated with growth traits in Chinese sheep breeds. In chicken, several CNV regions affecting *KCNJ11*, *MyoD1,* and *SOX6* were linked to the performance traits in boilers (Fernandes et al. 2021). In swine, various array platforms have been used to identify the CNVs, including comparative genomic hybridization arrays (Fadista et al. 2010; Li et al. 2012; Wang et al. 2014; Wang et al. 2015), Illumina Infinium II Multisample SNP assay (Wang et al. 2013; Long et al. 2016), and llumina PorcineSNP60 BeadChip (Schiavo et al. 2014; Ramayo-Caldas et al. 2010; Xie et al. 2016; Zhou et al. 2016; Hay et al. 2017). Relatively, a smaller number of studies used next-generation sequencing data to identify CNVs in the swine. One recent study used next-generation sequencing data to identify CNVs impacting genes important for the fatty acid composition and growth traits in an Iberian x Landrace backcross population (Revilla et al. 2017). Keel and coauthors reported more than 3000 CNVs in the porcine genome using 24 of the purebred founding boars. And more than half (1820 of them) of the CNVs had not been reported in previous studies (Keel et al. 2019).

5.5 CNVs and the Associated Diseases in Livestock Species

Many genes belonging to the innate and immune systems are affected by copy numbers in the genome of vertebrate species, with the genes in the major histocompatibility complex (MHC) as the most prominent ones (Balakrishnan et al. 2010; Ohtsuka et al. 2008). In cattle, significant enrichment of immune genes in the CNV regions was reported (Liu et al. 2009; Choi et al. 2013; Bickhart et al. 2012).

The immune-related genes affected by CNVs in the livestock species belong mainly to four categories: MHC, antimicrobial peptides (AMP), T cell receptors, and endogenous retrovirus. In the chickens, many CNVs appear to render varied susceptibility to Marek's disease (Bai et al. 2020; Luo et al. 2013). Notably, in the study by Luo and coauthors, a total of 45 CNVs were found in four lines of chicken, and 28 of these were involved in immune response and cell proliferation (Luo et al. 2013). In cattle, a trans-activator of MHC II, CIITA, was duplicated in cattle with resistance to gastrointestinal nematodes resistance and susceptibility (Liu et al. 2011). Additionally, cattle specific, β-defensin, antimicrobial peptides were found with varying copy numbers (Bickhart et al. 2012). β-defensins are typically produced by the epithelial tissues to defend against pathogen infection (reviewed in Weinberg, et al. (2012)). This gene family was reported with intraspecies copy number fluctuations (Liu et al. 2009, 2010; Bickhart et al. 2012). An AMP-coding gene, *BSP30A*, was found highly copy number variable among different cattle breeds (Bickhart et al. 2012). A cathelicidin-type AMP, *CATHL4*, had two to four fold increase in gene expression in Nelore cattle neutrophils (Flores 2011), owing to its elevated copy number. The workshop class I gene family (*WC1*), which encodes pattern recognition receptors expressed on γδT cells (Herzig and Baldwin 2009), is unique to the cattle, sheep, and swine genomes. Copy number variation in *WC1* has been reported in cattle (Liu et al. 2010; Bickhart et al. 2012; Chen et al. 2012). In sheep, an endogenous retrovirus, oncogenic exogenous Jaagsiekte sheep retrovirus (JSRV), is responsible for ovine pulmonary adenocarcinoma. It has several endogenous counterparts, termed enJSRV, which are highly duplicated in the sheep genome (Sistiaga-Poveda and Jugo 2014). CNV deletions that fall within the coding regions of the genome may have direct causal impact to the phenotype. Drögemüller and coauthors reported the effect of a partial deletion in the bovine ectodysplasin 1in gene to the Anhidrotic ectodermal dysplasia, which is characterized by reduced number of sweat glands and incisior anodontia in cattle (Drogemuller et al. 2001).

5.6 The Obstacles in Functional Annotation of CNVs and their Potential Application

The perceived functional role of CNVs is through gene dosage imbalances, disruptions of genes, and their regulatory elements. A typical genome-wide survey of CNVs yields a minimum of few hundred potential functional CNVs. Accurate functional annotation of the plethora of CNVs in the genome remains a major obstacle in the application of CNVs as functional biomarkers. In human and mouse models, CNVs have been associated with a number of genetic disorders, including autism spectrum disorders, autoimmune diseases, and neurodevelopmental disorders (Vorstman and Scherer 2021; Stamouli et al. 2018; Vicari et al. 2019; Larson et al. 2018; Barbosa et al. 2018; Pereira et al. 2019). With ever increasing amount of whole genome sequencing data generated by high-

throughput technologies, researchers must continuously improve bioinformatic softwares and annotation tools to help identify functionally relevant CNVs. According to the standard published by American College of Medical Genetics and Genomics (ACMG), the functional role of each CNV can be classified into these categories: benign, likely benign, a variant of uncertain significance, likely pathogenic, and pathogenic (Riggs et al. 2020). A few computational tools have been developed to help determine clinically important CNVs (Geoffroy et al. 2018; Ganel et al. 2017; Erikson et al. 2015), though each of them utilized different criteria in the determination of functional CNVs compared to that recommended by ACMG (Riggs et al. 2020). Recently, Gurbich and Illinsky (Gurbich and Ilinsky 2020) employed a scoring rubric that closely follow the standard recommended by ACMG. This tool considers the loss and gain of the CNV, the total number of protein-coding genes impacted by the targeted CNV, the degree of overlap (e.g., partial or complete overlap, the position of the exon involved) with established, haploinsufficient genes and genomic regions, and the population frequency of the CNV. Points were aggregated for each evaluated section of the rubric, which will determine the functional class of the targeted CNVs.

So far, more than 150 computational tools have been developed to identify structural variants using next-generation sequencing data (reviewed in Geoffroy et al. (Geoffroy et al. 2018)). These widely available bioinformatics tools undoubtedly help propel the research into CNVs and their potential functional impact in disease and health. However, each of the computation tools uses different algorithms and carries significant false positive error rate (Samarakoon et al. 2016). Furthermore, many newly identified CNVs are novel, requiring further investigations to determine their clinical and functional implications. Thus, manual annotation of the functional role remains a formidable task.

Besides the discrepancy in the computational tools in the functional annotation, other challenges persist. These include limited phenotypic data and conflicting associations among different studies to name a few. So far, most of the work in the functional annotation of CNVs relies on the common variants (Han et al. 2020; Consortium GTEx et al. 2017), mainly due to its sufficient statistical power to allow the association analysis between the population frequency of CNVs and phenotype of interest (Wellcome Trust Case Control Consortium et al. 2010). Impeded by the challenge in acquiring enough samples, the impact of rare CNVs remains largely untouched. To identify confident, actionable CNVs that can be used for livestock traits selection, well-curated database and gold-standard populations need to be established.

5.7 Future Directions

SNP microarrays and short reads generated by the next-generation sequencing have been primarily used to identify CNVs and in the subsequent analysis of their association with important production traits in livestock species. The limitations

exist in each of this technology. SNP microarray-based methods are known to have low probe density, low resolution, and significant hybridization noise. Additionally, array-based methods are confined by the predetermined SNP panel, thus offering no ability for de novo CNV identification in the regions not covered by the predesigned SNP array panel. With the short-read-based method, sequencing biases, repetitive genomic features, genomic polymorphisms, and translocation all make it difficult to assemble a substantial number of regions in the genome. To develop the CNVs as an efficient genomic improvement tool like the ones done in SNP-based genomic prediction, confident identification and annotation of causal CNVs are required. To achieve this, the first step is the development and routine update of CNV maps in various livestock species. So far, CNV maps have been developed for cattle (Butty et al. 2020; Liu et al. 2009; Zimin et al. 2012), chicken (Wang et al. 2010), sheep (Salehian-Dehkordi et al. 2021; Liu et al. 2013), and pig (Groenen et al. 2012) genomes. Given the prevalence of CNVs in mammalian genome and many gaped and unresolved regions in the genomes, these mapping efforts only stand a starting effort; the sequencing technology advancement and new development of computational tools will enable the fine mapping and routine updates of CNVs in the genome of many livestock species.

Most CNV detection algorithms rely on the comparison to a reference genome assembly, making the quality and completeness of a reference genome an indispensable requirement. The long reads generated by Pacific Biosciences (PacBio) can help close the assembly gap and correct assembly errors previously generated by short reads (English et al. 2012). In line with this, new methods and algorithms are needed for de novo assembly of the reference genome using the long reads. Traditional methods for sequence data analysis, like Burrows-Wheeler transform (BWT) (Burrows and Wheller 1994) and space-efficient graph representations with succinct de Bruijn graphs (Chikhi et al. 2019), have a trade-off between efficiency and information loss. A recently developed tool, mcBG, makes use of minimizer-space de Bruijn graphs to facilitate long-read genome assembly (Ekim et al. 2021). With the new sequencing technologies represented by the PacBio and Oxford Nanopore, genome assembly and annotation continue to improve from the longer reads with low error rates (Wenger et al. 2019). Even with short reads, improved algorithm and analytical methods can lead to new discoveries as demonstrated by the large-scale data reanalysis initiative (Edgar et al. 2022; Lachmann et al. 2018).

Most common CNV callers that use sequencing data for the identification of CNVs generally rely on the read-depth, read-pair, and split-reads to identify relatively simple CNVs (Alkan et al. 2011; Medvedev and Brudno 2008), e.g., duplication and deletions. Notably, read depth-based methods can detect the presence or absence of CNVs, but lack the ability to determine the location of such CNV variations. By integrate multiple sequencing signatures, modern CNV callers, such as DELLY (Rausch et al. 2012) LUMPY (Layer et al. 2014), SV-Bay (Iakovishina et al. 2016), TIDDIT (Eisfeldt et al. 2017), SVelter (Zhao et al. 2016) and TARDIS (Soylev et al. 2019), expand the spectrum of CNV identification to novel insertions, inversions, and mobile element insertions. There is still a need to method development for the accurate characterization of complex CNVs that are tandem or interspersed segmental duplications (Soylev et al. 2019; Chaisson et al. 2019).

References

Alkan C, Coe BP, Eichler EE (2011) Genome structural variation discovery and genotyping. Nat Rev Genet 12(5):363–376

Almal SH, Padh H (2012) Implications of gene copy-number variation in health and diseases. J Hum Genet 57(1):6–13

Bai H, He Y, Ding Y, Chu Q, Lian L, Heifetz EM, Yang N, Cheng HH, Zhang H, Chen J et al (2020) Genome-wide characterization of copy number variations in the host genome in genetic resistance to Marek's disease using next generation sequencing. BMC Genet 21(1):77

Balakrishnan CN, Ekblom R, Volker M, Westerdahl H, Godinez R, Kotkiewicz H, Burt DW, Graves T, Griffin DK, Warren WC et al (2010) Gene duplication and fragmentation in the zebra finch major histocompatibility complex. BMC Biol 8:29

Barbosa FB, Simioni M, Wiezel CEV, Torres FR, Molck MC, Bonilla MM, de Araujo TK, Donadi EA, Gil-da-Silva-Lopes VL, Lemos B et al (2018) Copy number variation in the susceptibility to systemic lupus erythematosus. PLoS One 13(11):e0206683

Bayes M, Magano LF, Rivera N, Flores R, Perez Jurado LA (2003) Mutational mechanisms of Williams-Beuren syndrome deletions. Am J Hum Genet 73(1):131–151

Belshaw R, Pereira V, Katzourakis A, Talbot G, Paces J, Burt A, Tristem M (2004) Long-term reinfection of the human genome by endogenous retroviruses. Proc Natl Acad Sci U S A 101(14):4894–4899

Bickhart DM, Liu GE (2014) The challenges and importance of structural variation detection in livestock. Front Genet 5:37

Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, Matukumalli LK, Song J, Schnabel RD, Ventura M, Taylor JF et al (2012) Copy number variation of individual cattle genomes using next-generation sequencing. Genome Res 22(4):778–790

Brenig B, Beck J, Floren C, Bornemann-Kolatzki K, Wiedemann I, Hennecke S, Swalve H, Schutz E (2013) Molecular genetics of coat colour variations in white Galloway and White Park cattle. Anim Genet 44(4):450–453

Burrows MW, Wheller DJ (1994) A block-sorting lossless data compression algorithm. Digital SRC Research Report Citeseer, p 124

Butty AM, Chud TCS, Miglior F, Schenkel FS, Kommadath A, Krivushin K, Grant JR, Hafliger IM, Drogemuller C, Canovas A et al (2020) High confidence copy number variants identified in Holstein dairy cattle from whole genome sequence and genotype array data. Sci Rep 10(1):8044

Chaisson MJP, Sanders AD, Zhao X, Malhotra A, Porubsky D, Rausch T, Gardner EJ, Rodriguez OL, Guo L, Collins RL et al (2019) Multi-platform discovery of haplotype-resolved structural variation in human genomes. Nat Commun 10(1):1784

Chen C, Herzig CT, Alexander LJ, Keele JW, McDaneld TG, Telfer JC, Baldwin CL (2012) Gene number determination and genetic polymorphism of the gamma delta T cell co-receptor WC1 genes. BMC Genet 13:86

Chen L, Zhou W, Zhang L, Zhang F (2014) Genome architecture and its roles in human copy number variation. Genomics Inform 12(4):136–144

Cheng J, Jiang R, Yang Y, Cao X, Huang Y, Lan X, Lei C, Hu L, Chen H (2020) Association analysis of KMT2D copy number variation as a positional candidate for growth traits. Gene 753: 144799

Chikhi R, Holub J, Medvedev P (2019) Data structures to represent sets of k-long DNA sequences. ACM Comput Surv 54(1):1–22

Choi JW, Lee KT, Liao X, Stothard P, An HS, Ahn S, Lee S, Lee SY, Moore SS, Kim TH (2013) Genome-wide copy number variation in Hanwoo, black Angus, and Holstein cattle. Mamm Genome 24(3–4):151–163

Clop A, Vidal O, Amills M (2012) Copy number variation in the genomes of domestic animals. Anim Genet 43(5):503–517

Consortium GTEx, Laboratory DA, Coordinating Center-Analysis Working G, Statistical Methods groups-Analysis Working Group, Enhancing GTEx groups, Fund NIH Common, NIH/NCI,

NIH/NHGRI, NIH/NIMH, NIH/NIDA et al (2017) Genetic effects on gene expression across human tissues. Nature 550(7675):204–213
Cook EH Jr, Scherer SW (2008) Copy-number variations associated with neuropsychiatric conditions. Nature 455(7215):919–923
Cumer T, Boyer F, Pompanon F (2021) Genome-wide detection of structural variations reveals new regions associated with domestication in small ruminants. Genome Biol Evol 13(8):evab165
de Almeida Santana MH, Junior GA, Cesar AS, Freua MC, da Costa GR, da Luz ESS, Leme PR, Fukumasu H, Carvalho ME, Ventura RV et al (2016) Copy number variations and genome-wide associations reveal putative genes and metabolic pathways involved with the feed conversion ratio in beef cattle. J Appl Genet 57(4):495–504
Dopman EB, Hartl DL (2007) A portrait of copy-number polymorphism in Drosophila melanogaster. Proc Natl Acad Sci U S A 104(50):19920–19925
Drogemuller C, Distl O, Leeb T (2001) Partial deletion of the bovine ED1 gene causes anhidrotic ectodermal dysplasia in cattle. Genome Res 11(10):1699–1705
Durkin K, Coppieters W, Drogemuller C, Ahariz N, Cambisano N, Druet T, Fasquelle C, Haile A, Horin P, Huang L et al (2012) Serial translocation by means of circular intermediates underlies colour sidedness in cattle. Nature 482(7383):81–84
Edgar RC, Taylor J, Lin V, Altman T, Barbera P, Meleshko D, Lohr D, Novakovsky G, Buchfink B, Al-Shayeb B et al (2022) Petabase-scale sequence alignment catalyses viral discovery. Nature 602(7895):142–147
Eisfeldt J, Vezzi F, Olason P, Nilsson D, Lindstrand A (2017) TIDDIT, an efficient and comprehensive structural variant caller for massive parallel sequencing data. F1000Res 6:664
Ekim B, Berger B, Chikhi R (2021) Minimizer-space de Bruijn graphs: whole-genome assembly of long reads in minutes on a personal computer. Cell Syst 12(10):958–968 e956
English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC et al (2012) Mind the gap: upgrading genomes with Pacific biosciences RS long-read sequencing technology. PLoS One 7(11):e47768
Erikson GA, Deshpande N, Kesavan BG, Torkamani A (2015) SG-ADVISER CNV: copy-number variant annotation and interpretation. Genet Med 17(9):714–718
Fadista J, Thomsen B, Holm LE, Bendixen C (2010) Copy number variation in the bovine genome. BMC Genomics 11:284
Fernandes AC, da Silva VH, Goes CP, Moreira GCM, Godoy TF, Ibelli AMG, Peixoto JO, Cantao ME, Ledur MC, de Rezende FM et al (2021) Genome-wide detection of CNVs and their association with performance traits in broilers. BMC Genomics 22(1):354
Flores EG (2011) Characterization of the bovine cathelicidin gene family. Doctoral dissertation. Texas A&M University
Frenkel S, Bernstein CN, Sargent M, Jiang W, Kuang Q, Xu W, Hu P (2020) Copy number variation-based gene set analysis reveals cytokine signalling pathways associated with psychiatric comorbidity in patients with inflammatory bowel disease. Genomics 112(1):683–693
Fu W, Zhang F, Wang Y, Gu X, Jin L (2010) Identification of copy number variation hotspots in human populations. Am J Hum Genet 87(4):494–504
Ganel L, Abel HJ, FinMetSeq C, Hall IM (2017) SVScore: an impact prediction tool for structural variation. Bioinformatics 33(7):1083–1085
Gao Y, Jiang J, Yang S, Hou Y, Liu GE, Zhang S, Zhang Q, Sun D (2017) CNV discovery for milk composition traits in dairy cattle using whole genome resequencing. BMC Genomics 18(1):265
Geistlinger L, da Silva VH, Cesar ASM, Tizioto PC, Waldron L, Zimmer R, Regitano LCA, Coutinho LL (2018) Widespread modulation of gene expression by copy number variation in skeletal muscle. Sci Rep 8(1):1399
Geoffroy V, Herenger Y, Kress A, Stoetzel C, Piton A, Dollfus H, Muller J (2018) AnnotSV: an integrated tool for structural variations annotation. Bioinformatics 34(20):3572–3574
Girirajan S, Campbell CD, Eichler EE (2011) Human copy number variation and complex genetic disease. Annu Rev Genet 45:203–226

Girirajan S, Dennis MY, Baker C, Malig M, Coe BP, Campbell CD, Mark K, Vu TH, Alkan C, Cheng Z et al (2013) Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. Am J Hum Genet 92(2):221–237

Gorla E, Cozzi MC, Roman-Ponce SI, Ruiz Lopez FJ, Vega-Murillo VE, Cerolini S, Bagnato A, Strillacci MG (2017) Genomic variability in Mexican chicken population using copy number variants. BMC Genet 18(1):61

Goshu HAXW, Chu M, Pengjia B, Xue Zhi D, Yan P (2020) Novel copy number variations of the CHRM3 gene associated with gene expression and growth traits in Chinese Datong yak (Bos grunniens). J Appl Anim Res 48:156–165

Groenen MA, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard-C, Park C, Milan D, Megens HJ et al (2012) Analyses of pig genomes provide insight into porcine demography and evolution. Nature 491(7424):393–398

Gurbich TA, Ilinsky VV (2020) ClassifyCNV: a tool for clinical annotation of copy-number variants. Sci Rep 10(1):20375

Han L, Zhao X, Benton ML, Perumal T, Collins RL, Hoffman GE, Johnson JS, Sloofman L, Wang HZ, Stone MR et al (2020) Functional annotation of rare structural variation in the human brain. Nat Commun 11(1):2990

Hastings PJ, Lupski JR, Rosenberg SM, Ira G (2009) Mechanisms of change in gene copy number. Nat Rev Genet 10(8):551–564

Hay EHA, Choi I, Xu L, Zhou Y, Rowland RRR, Lunney JK, Liu GE (2017) CNV analysis of host responses to porcine reproductive and respiratory syndrome virus infection. J Genomics 5:58–63

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92(2):433–443

Henrichsen CN, Vinckenbosch N, Zollner S, Chaignat E, Pradervand S, Schutz F, Ruedi M, Kaessmann H, Reymond A (2009a) Segmental copy number variation shapes tissue transcriptomes. Nat Genet 41(4):424–429

Henrichsen CN, Chaignat E, Reymond A (2009b) Copy number variants, diseases and gene expression. Hum Mol Genet 18(R1):R1–R8

Herzig CT, Baldwin CL (2009) Genomic organization and classification of the bovine WC1 genes and expression by peripheral blood gamma delta T cells. BMC Genomics 10:191

Hou Y, Liu GE, Bickhart DM, Cardone MF, Wang K, Kim ES, Matukumalli LK, Ventura M, Song J, VanRaden PM et al (2011) Genomic characteristics of cattle copy number variations. BMC Genomics 12:127

Huang C, Ge F, Ren W, Zhang Y, Wu X, Zhang Q, Ma X, Bao P, Guo X, Chu M et al (2021) Copy number variation of the HPGDS gene in the Ashidan yak and its associations with growth traits. Gene 772:145382

Iakovishina D, Janoueix-Lerosey I, Barillot E, Regnier M, Boeva V (2016) SV-bay: structural variant detection in cancer genomes using a Bayesian approach with correction for GC-content and read mappability. Bioinformatics 32(7):984–992

Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet 11(2):97–108

Iskow RC, Gokcumen O, Lee C (2012) Exploring the role of copy number variants in human adaptation. Trends Genet 28(6):245–257

Kang X, Li M, Liu M, Liu S, Pan MG, Wiggans GR, Rosen BD, Liu GE (2020) Copy number variation analysis reveals variants associated with milk production traits in dairy goats. Genomics 112(6):4934–4937

Keel BN, Nonneman DJ, Lindholm-Perry AK, Oliver WT, Rohrer GA (2019) A survey of copy number variation in the porcine genome detected from whole-genome sequence. Front Genet 10:737

Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, Silverstein MC, Ma'ayan A (2018) Massive mining of publicly available RNA-seq data from human and mouse. Nat Commun 9(1):1366

Larson FV, Arrand JR, Tantam D, Jones PB, Holland AJ (2018) Copy number variants in people with autism spectrum disorders and co-morbid psychosis. Eur J Med Genet 61(4):230–234

Layer RM, Chiang C, Quinlan AR, Hall IM (2014) LUMPY: a probabilistic framework for structural variant discovery. Genome Biol 15(6):R84

Li Y, Mei S, Zhang X, Peng X, Liu G, Tao H, Wu H, Jiang S, Xiong Y, Li F (2012) Identification of genome-wide copy number variations among diverse pig breeds by array CGH. BMC Genomics 13:725

Li B, Zhang N, Wang YG, George AW, Reverter A, Li Y (2018) Genomic prediction of breeding values using a subset of SNPs identified by three machine learning methods. Front Genet 9:237

Liu GE, Ventura M, Cellamare A, Chen L, Cheng Z, Zhu B, Li C, Song J, Eichler EE (2009) Analysis of recent segmental duplications in the bovine genome. BMC Genomics 10:571

Liu GE, Hou Y, Zhu B, Cardone MF, Jiang L, Cellamare A, Mitra A, Alexander LJ, Coutinho LL, Dell'Aquila ME et al (2010) Analysis of copy number variations among diverse cattle breeds. Genome Res 20(5):693–703

Liu GE, Brown T, Hebert DA, Cardone MF, Hou Y, Choudhary RK, Shaffer J, Amazu C, Connor EE, Ventura M et al (2011) Initial analysis of copy number variations in cattle selected for resistance or susceptibility to intestinal nematodes. Mamm Genome 22(1–2):111–121

Liu J, Zhang L, Xu L, Ren H, Lu J, Zhang X, Zhang S, Zhou X, Wei C, Zhao F et al (2013) Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. BMC Genomics 14:229

Liu M, Lia B, Huanga Y, Yang M, Lana X, Lei C, Qub W, Bai Y, Chena H (2016) Copy number variation of bovine MAPK10 modulates the transcriptional activity and affects growth traits. Livest Sci 194:44–50. https://doi.org/10.1016/j.livsci.2016.09.014

Liu M, Fang L, Liu S, Pan MG, Seroussi E, Cole JB, Ma L, Chen H, Liu GE (2019) Array CGH-based detection of CNV regions and their potential association with reproduction and other economic traits in Holsteins. BMC Genomics 20(1):181

Long Y, Su Y, Ai H, Zhang Z, Yang B, Ruan G, Xiao S, Liao X, Ren J, Huang L et al (2016) A genome-wide association study of copy number variations with umbilical hernia in swine. Anim Genet 47(3):298–305

Luo J, Yu Y, Mitra A, Chang S, Zhang H, Liu G, Yang N, Song J (2013) Genome-wide copy number variant analysis in inbred chickens lines with different susceptibility to Marek's disease. G3 (Bethesda) 3(2):217–223

Lupski JR (2007) Genomic rearrangements and sporadic disease. Nat Genet 39(7 Suppl):S43–S47

Medvedev P, Brudno M (2008) Ab initio whole genome shotgun assembly with mated short reads. In: Annual International Conference on Research in Computational Molecular Biology. Springer, Berlin, Heidelberg, pp 50–64

Memon D, Gill MB, Papachristou EK, Ochoa D, D'Santos CS, Miller ML, Beltrao P (2021) Copy number aberrations drive kinase rewiring, leading to genetic vulnerabilities in cancer. Cell Rep 35(7):109155

Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK et al (2011) Mapping copy number variation by population-scale genome sequencing. Nature 470(7332):59–65

Moasser MM (2007) The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. Oncogene 26(45):6469–6487

Norris BJ, Whan VA (2008) A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. Genome Res 18(8):1282–1293

Ohtsuka M, Inoko H, Kulski JK, Yoshimura S (2008) Major histocompatibility complex (Mhc) class Ib gene duplications, organization and expression patterns in mouse strain C57BL/6. BMC Genomics 9:178

Pereira KMC, Perazzio S, Faria AGA, Moreira ES, Santos VC, Grecco M, da Silva NP, Andrade LEC (2019) Impact of C4, C4A and C4B gene copy number variation in the susceptibility, phenotype and progression of systemic lupus erythematosus. Adv Rheumatol 59(1):36

Pierce MD, Dzama K, Muchadeyi FC (2018) Genetic diversity of seven cattle breeds inferred using copy number variations. Front Genet 9:163

Rafter P, Purfield DC, Berry DP, Parnell AC, Gormley IC, Kearney JF, Coffey MP, Carthy TR (2018) Characterization of copy number variants in a large multibreed population of beef and dairy cattle using high-density single nucleotide polymorphism genotype data. J Anim Sci 96(10):4112–4124

Ramayo-Caldas Y, Castello A, Pena RN, Alves E, Mercade A, Souza CA, Fernandez AI, Perez-Enciso M, Folch JM (2010) Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip. BMC Genomics 11:593

Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO (2012) DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 28(18):i333–i339

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W et al (2006) Global variation in copy number in the human genome. Nature 444(7118):444–454

Revilla M, Puig-Oliveras A, Castello A, Crespo-Piazuelo D, Paludo E, Fernandez AI, Ballester M, Folch JM (2017) A global analysis of CNVs in swine using whole genome sequence data and association analysis with fatty acid composition and growth traits. PLoS One 12(5):e0177014

Rigau M, Juan D, Valencia A, Rico D (2019) Intronic CNVs and gene expression variation in human populations. PLoS Genet 15(1):e1007902

Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, Raca G, Ritter DI, South ST, Thorland EC et al (2020) Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the clinical genome resource (ClinGen). Genet Med 22(2):245–257

Salehian-Dehkordi H, Xu YX, Xu SS, Li X, Luo LY, Liu YJ, Wang DF, Cao YH, Shen M, Gao L et al (2021) Genome-wide detection of copy number variations and their association with distinct phenotypes in the World's sheep. Front Genet 12:670582

Samarakoon PS, Sorte HS, Stray-Pedersen A, Rodningen OK, Rognes T, Lyle R (2016) cnvScan: a CNV screening and annotation tool to improve the clinical utility of computational CNV prediction from exome sequencing data. BMC Genomics 17:17

Sasaki S, Watanabe T, Nishimura S, Sugimoto Y (2016) Genome-wide identification of copy number variation using high-density single-nucleotide polymorphism array in Japanese black cattle. BMC Genet 17:26

Schaeffeler E, Schwab M, Eichelbaum M, Zanger UM (2003) CYP2D6 genotyping strategy based on gene copy number determination by TaqMan real-time PCR. Hum Mutat 22(6):476–485

Schiavo G, Dolezal MA, Scotti E, Bertolini F, Calo DG, Galimberti G, Russo V, Fontanesi L (2014) Copy number variants in Italian large white pigs detected using high-density single nucleotide polymorphisms and their association with back fat thickness. Anim Genet 45(5):745–749

Schuster-Bockler B, Conrad D, Bateman A (2010) Dosage sensitivity shapes the evolution of copy-number varied regions. PLoS One 5(3):e9474

Shi T, Xu Y, Yang M, Huang Y, Lan X, Lei C, Qi X, Yang X, Chen H (2016) Copy number variations at LEPR gene locus associated with gene expression and phenotypic traits in Chinese cattle. Anim Sci J 87(3):336–343

Shwan NAA, Louzada S, Yang F, Armour JAL (2017) Recurrent rearrangements of human amylase genes create multiple independent CNV series. Hum Mutat 38(5):532–539

Silva VH, Regitano LC, Geistlinger L, Pertille F, Giachetto PF, Brassaloti RA, Morosini NS, Zimmer R, Coutinho LL (2016) Genome-wide detection of CNVs and their association with meat tenderness in Nelore cattle. PLoS One 11(6):e0157711

Sistiaga-Poveda M, Jugo BM (2014) Evolutionary dynamics of endogenous Jaagsiekte sheep retroviruses proliferation in the domestic sheep, mouflon and Pyrenean chamois. Heredity (Edinb) 112(6):571–578

Soylev A, Le TM, Amini H, Alkan C, Hormozdiari F (2019) Discovery of tandem and interspersed segmental duplications using high-throughput sequencing. Bioinformatics 35(20):3923–3930

Stamouli S, Anderlid BM, Willfors C, Thiruvahindrapuram B, Wei J, Berggren S, Nordgren A, Scherer SW, Lichtenstein P, Tammimies K et al (2018) Copy number variation analysis of 100 twin pairs enriched for neurodevelopmental disorders. Twin Res Hum Genet 21(1):1–11

Stankiewicz P, Lupski JR (2010) Structural variation in the human genome and its role in disease. Annu Rev Med 61:437–455

Toremurat Z, Ibrahim EE, Huang YZ, Lan X, Pi L, Chaogetu B, Hu L, Chen H (2022) Copy number variations of TOP2B gene are associated with growth traits in Chinese sheep breeds. Anim Biotechnol 33(1):85–89

Upadhyay M, da Silva VH, Megens HJ, Visker M, Ajmone-Marsan P, Balteanu VA, Dunner S, Garcia JF, Ginja C, Kantanen J et al (2017) Distribution and functionality of copy number variation across European cattle populations. Front Genet 8:108

VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS (2009) Invited review: reliability of genomic predictions for north American Holstein bulls. J Dairy Sci 92(1):16–24

Vicari S, Napoli E, Cordeddu V, Menghini D, Alesi V, Loddo S, Novelli A, Tartaglia M (2019) Copy number variants in autism spectrum disorders. Prog Neuro-Psychopharmacol Biol Psychiatry 92:421–427

Viginier B, Dolmazon C, Lantier I, Lantier F, Archer F, Leroux C, Terzian C (2012) Copy number variation and differential expression of a protective endogenous retrovirus in sheep. PLoS One 7(7):e41965

Vorstman J, Scherer SW (2021) What a finding of gene copy number variation can add to the diagnosis of developmental neuropsychiatric disorders. Curr Opin Genet Dev 68:18–25

Wang X, Nahashon S, Feaster TK, Bohannon-Stewart A, Adefope N (2010) An initial map of chromosomal segmental copy number variations in the chicken. BMC Genomics 11:351

Wang J, Wang H, Jiang J, Kang H, Feng X, Zhang Q, Liu JF (2013) Identification of genome-wide copy number variations among diverse pig breeds using SNP genotyping arrays. PLoS One 8(7):e68683

Wang J, Jiang J, Wang H, Kang H, Zhang Q, Liu JF (2014) Enhancing genome-wide copy number variation identification by high density array CGH using diverse resources of pig breeds. PLoS One 9(1):e87571

Wang J, Jiang J, Wang H, Kang H, Zhang Q, Liu JF (2015) Improved detection and characterization of copy number variations among diverse pig breeds by array CGH. G3 (Bethesda) 5(6): 1253–1261

Wang Z, Guo J, Guo Y, Yang Y, Teng T, Yu Q, Wang T, Zhou M, Zhu Q, Wang W et al (2020) Genome-wide detection of CNVs and association with body weight in sheep based on 600K SNP arrays. Front Genet 11:558

Weinberg A, Jin G, Sieg S, McCormick TS (2012) The yin and yang of human Beta-defensins in health and disease. Front Immunol 3:294

Wellcome Trust Case Control Consortium, Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, Vukcevic D, Barnes C, Conrad DF et al (2010) Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature 464(7289):713–720

Wenger AM, Peluso P, Rowell WJ, Chang PC, Hall RJ, Concepcion GT, Ebler J, Fungtammasan A, Kolesnikov A, Olson ND et al (2019) Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat Biotechnol 37(10): 1155–1162

Xie J, Li R, Li S, Ran X, Wang J, Jiang J, Zhao P (2016) Identification of copy number variations in Xiang and Kele pigs. PLoS One 11(2):e0148565

Xu Y, Shi T, Cai H, Zhou Y, Lan X, Zhang C, Lei C, Qi X, Chen H (2014a) Associations of MYH3 gene copy number variations with transcriptional expression and growth traits in Chinese cattle. Gene 535(2):106–111

Xu L, Cole JB, Bickhart DM, Hou Y, Song J, VanRaden PM, Sonstegard TS, Van Tassell CP, Liu GE (2014b) Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. BMC Genomics 15:683

Xu Y, Jiang Y, Shi T, Cai H, Lan X, Zhao X, Plath M, Chen H (2017) Whole-genome sequencing reveals mutational landscape underlying phenotypic differences between two widespread Chinese cattle breeds. PLoS One 12(8):e0183921

Young JM, Endicott RM, Parghi SS, Walker M, Kidd JM, Trask BJ (2008) Extensive copy-number variation of the human olfactory receptor gene family. Am J Hum Genet 83(2):228–242

Zarrei M, MacDonald JR, Merico D, Scherer SW (2015) A copy number variation map of the human genome. Nat Rev Genet 16(3):172–183

Zhang X, Wang K, Wang L, Yang Y, Ni Z, Xie X, Shao X, Han J, Wan D, Qiu Q (2016) Genome-wide patterns of copy number variation in the Chinese yak genome. BMC Genomics 17:379

Zhang GM, Zheng L, He H, Song CC, Zhang ZJ, Cao XK, Lei CZ, Lan XY, Qi XL, Chen H et al (2018) Associations of GBP2 gene copy number variations with growth traits and transcriptional expression in Chinese cattle. Gene 647:101–106

Zhang Y, Hu Y, Wang X, Jiang Q, Zhao H, Wang J, Ju Z, Yang L, Gao Y, Wei X et al (2019) Population structure, and selection signatures underlying high-altitude adaptation inferred from genome-wide copy number variations in Chinese indigenous cattle. Front Genet 10:1404

Zhao X, Emery SB, Myers B, Kidd JM, Mills RE (2016) Resolving complex structural genomic rearrangements using a randomized approach. Genome Biol 17(1):126

Zhou LS, Li J, Yang J, Liu CL, Xie XH, He YN, Liu XX, Xin WS, Zhang WC, Ren J et al (2016) Genome-wide mapping of copy number variations in commercial hybrid pigs using a high-density Snp genotyping Array. Genetika 52(1):97–105

Zhou Y, Connor EE, Wiggans GR, Lu Y, Tempelman RJ, Schroeder SG, Chen H, Liu GE (2018) Genome-wide copy number variant analysis reveals variants associated with 10 diverse production traits in Holstein cattle. BMC Genomics 19(1):314

Zimin AV, Kelley DR, Roberts M, Marcais G, Salzberg SL, Yorke JA (2012) Mis-assembled "segmental duplications" in two versions of the Bos taurus genome. PLoS One 7(8):e42680

Chapter 6
Implementation of Genomic Selection under the Indian Scenario through Utilizing the Vast Repository of Female Datasets

Supriya Chhotaray and Vikas Vohra

Abstract Improvement of economically important quantitative traits in livestock has relied on BLUP estimation of breeding values based on phenotypes and pedigree records, through progeny testing programs. However, with advanced next-generation sequencing technology, during the last decade, the use of genome-wide markers for genomic evaluation of breeding values for dairy Sires has been the choice for genetic evaluation. Genomic selection along with conventional progeny testing programs has now proved to have a higher rate of genetic gains in livestock production and management systems. India although has successfully implemented the progeny testing programs to improve the rate of genetic gain, despite the large available genetic resources, efficient implementation of the genomic selection is still lagging. This chapter summarizes the genomic evaluation methods, opportunity for implementation of genetic selection in the Indian conditions, challenges, and various prospects of genomic selection.

Keywords Livestock · Breeding value · Genomic selection · Indian condition

6.1 Introduction

The discovery of the genome has been the key landmark in the journey of animal improvement via conventional breeding methods to genomic selection in the last decades. A set of contemporary genome analysis technologies is implemented worldwide as selection tools in dairy cattle breeding programs. Genomic information is also used largely for adjusted mating, validation of kinship, and conservation strategies for the sustainable production of dairy cattle. The implementation of genomic information in intensive progeny testing programs for production and reproductive traits has been a crucial step in increasing trends of dairy cattle productivity. From the perspective of the breeding objective, the principal effect of

S. Chhotaray · V. Vohra (✉)
Animal Genetics and Breeding Division, ICAR—National Dairy Research Institute, Karnal, Haryana, India

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_6

genomic selection is increased reliabilities of the breeding values used in the index. The genomic selection also renders lower costs of producing proven bulls, greater rates of genetic gain from shorter generation intervals, detection of previously unknown genetic disorders, and also identification of genes that influence economically important traits.

A sustainable dairy production system depends on the methods used to select the Sire (male) and Dam (female) to be parents of next-generation in an organized dairy herd as superior genetics lead to superior production. This mainly depends on methods used to select the Sires, as genetically a Sire is considered as more than half of the herd. To date, various methods of dairy animal selection have been used, which were largely the indices of economic traits-based selection of female animals. Selection of dairy Sires are progeny-based or based on the performance of the bulls' daughter, whereas males are selected for dairy traits based on the performance of their daughters. Improved animal genetics and breeding methods have been largely focused on the selection of superior Sire. In agrarian countries like India, and unlike the developed countries, the small size of animal holdings in a dairy herd renders any selection strategy inefficient and difficult to implement, yet considerable genetic improvement has been achieved using conventional methods of dairy Sire selection. In absence of a defined breeding plan and well-defined national genetic evaluation system, the breeding decision of the dairy Sire depends solely on the will of an individual livestock breeder/farmer.

In dairy cattle, most of the important economic traits are expressed only in females; hence, historically evaluation and characterization of Sire were being done based on the daughters' performance records. This is called progeny testing; and it has been the sole way to accurately determine *estimated breeding values (EBV)* based on phenotypic information (traits measured on an individual, like milk yield, and body height) and to select Sires based on their progeny performance. Sattler (2013) described progeny testing as a three-fold program that provides information on Sire selection for artificial insemination (AI), data needed for effective genetic evaluations, and generates genetically proven young bulls for wide-scale use in AI programs. However, the technique has its own disadvantages that the selection decisions in a breeding program are postponed in the progeny testing (PT) program and the generation interval is lengthened. Evaluation through PT programs is relatively slow since considerable time is required to gather adequate daughter records for genetic evaluations of bulls with high accuracy. There comes the role of genomic selection in the genetic improvement of dairy sires. In conventional PT programs, the generation intervals of Sires and Dams of bulls are 7 and 4 years, respectively, while it is significantly reduced to approximately 2.5 years for both the Sires and Dams of bulls with the use of genomic markers in genetic evaluations (García-Ruiz et al. 2016). The generation interval refers to the average age of the Sire at which it becomes a parent; in a breeding program no Sire can be used for breeding the cows unless its genetic merit is proven, thus, to prove a Sire, i.e., by the time the daughter of a Sire performs a large time lapse thereby increasing the generation interval, but under genomic selection strategy, the time required to select the Sire is considerably low, as low as age of sexual maturity in bulls.

Genomic breeding value (GBV) can be estimated in a dairy herd as early as the day an animal is born.

As extensive pedigree records are not required to estimate genomic breeding values without much loss of accuracy, in the last few decades it has been reckoned that genomic selection can offer great potential in low- to medium-income countries like India. However, to successfully implement the genomic selection program, there remain several barriers such as lack of an effective recording system, difficulty in forming an appropriate reference population, sustainability of the livestock resource, lack of on-farm infrastructure, lack of skilled workers in the laboratory, and minimal computing facilities. This paper examines those challenges and explores opportunities to mitigate or reduce the problems, intending to enable smallholder livestock-keepers and their associated value chains in low- to middle-income countries to also benefit directly from the genomic selection. The present chapter is an attempt to comprehend the implementation of genetic selection in Indian dairying using the unorganized but vast repository of female data available in India.

6.2 Development of Genomic Selection

Several conventional dairy Sire selection methods discussed in the previous section have played a major role in the improvement of dairy production. The Best Linear Unbiased Prediction (BLUP) method of genetic evaluation eventually became the method of choice worldwide for the evaluation and selection of Sires and was rapidly incorporated into the methodology of progeny testing (PT) programs.

All of these genetic evaluation methods have assumed the Infinitesimal Model described by Fisher (1918) which assumes that production performance is controlled by an infinite number of loci (genes) and each of them adds an infinitesimally small quantity to the total genetic worth of an individual. The genetic merit of an animal is estimated as the sumtotal of the effects exerted by the infinite number of loci as the correct number of loci that were involved and their effect size hasn't been fully determined yet. Scientists are continually working on the discovery of loci with large effects and the actual number of loci involved in the expression of a phenotype immediately after the molecular structure of DNA, the genetic material, was determined by Watson and Crick (Crick and Watson 1953).

Improvement of economically important quantitative traits in Livestock has relied on BLUP estimation of 'Breeding Values' based on phenotypes and pedigree records. Through this strategy, average milk yield has increased by 5000 kg in the past 40 years in the American dairy cattle population of which 60% is due to genetic improvement (Zhang et al. 2011). However, traditional selection methods faced limitations in terms of efficiency, particularly for the traits which are lowly heritable, longevity, sex-limited, and carcass traits. To overcome these limitations, it was assumed that the discovery of useful genetic markers may tackle the situation.

During the post-genomic era, the advancements in the field of molecular genetics have led to extensive use of genetic markers in the past 20 years for genetic

improvement through marker-assisted selection (MAS). The accuracy of selection is likely to increase by MAS programs. However, the application of MAS in animal breeding practices has been very limited due to several reasons. Molecular geneticists used markers that were large DNA segments (100–1000 base pairs), but were not very close to the actual genes. As the success of MAS depends on the closeness of marker and genes, the term linkage disequilibrium (LD) was used to denote the closeness. High LD indicates more closeness of the marker to the gene and fewer chances of recombination between the marker allele and gene allele across generations. MAS requires LD of 30% or more between markers and genes, to act as markers to the genes. In practice, very few markers have some significant effects on the economic traits and a marginal proportion of genetic variation is explained by each of these markers. Determination of marker genotypes often lacks consistency; consequently, the selection applied on marker genotypes was not as successful as expected. Determination of DNA markers mostly relied on restriction fragment length polymorphisms (RFLP); however, the discovery of microsatellites that were much smaller segments of DNA proved to be a better marker, but only a few microsatellites could be found for quantitative traits in Livestock.

In 1984, the Human Genome Project was conceived with the aim to sequence all of DNA and was finally published in April 2003. Scientists were able to identify gene sequences and finally concluded that approximately 25,000 genes exist in the human genome and discovered single nucleotide polymorphisms (SNP) by comparing genome sequences of different individuals. The genome was found to be full of millions of SNPs. By 2004, an SNP panel containing 5000 SNP at one time was developed to genotype dairy bull or cow. Genotyping of dairy bulls in Canada was initially started with a 5000 SNP panel, known as the 5 K chip. Meuwissen et al. (2001) through a simulation of thousands of evenly spaced SNPs across the genome could estimate each SNP allele's effect on the target trait. Furthermore, those estimates could be applied to a different herd of bulls that have not been evaluated yet have genotype information, but not any daughter performance record. Hence, they showed that in the simulations, the breeding value could be estimated with an accuracy of 0.81, which is much larger than the accuracy of EBV obtained through a simple parent average that is generally less than 0.40. Schaeffer (2006) outlined the advantages of genomic estimation of breeding value and how this strategy could double the genetic progress in dairy cattle by saving 5 years of progeny testing a bull.

The first genomic studies in Canada involving less than 500 proven (PT) bulls and 10 K SNP chip revealed that in the validation group accuracy was 0.50 to 0.60 instead of 0.81 as in Meuwissen et al. (2001). The genetic gain was due to the selection decision made on young bulls 5 years earlier than the ongoing progeny test schemes and not from the increased accuracy of the Genomic Estimated Breeding Value (GEBV). Around the year 2006, several thousand dairy bulls from various geographical regions were genotyped through joint ventures of different countries and USDA. The project also involved the development of a new 50 K chip. More than 3000 bulls were genotyped during this project and calculated GEBV were made official in the United States and Canada.

6.2.1 The Process of Genomic Prediction

The Human Genome Project when published the first draft of the genome sequence, the use of biallelic markers called SNPs (pronounced as 'Snips') came into the limelight and its use in conventional BLUP models led to the development of genomic selection. To date, several SNP chip panels have been developed using various methods of next-generation sequencing to identify new SNPs, the most recent one being the deep sequencing methods. SNPs are the most common genetic variants used today in the genomic-based prediction of performance in Livestock and dairy Sire evaluation in a selection program (k/a genomic selection).

The major success of genomic selection lies in the estimation of individual SNP allele effects on the trait of economic interest in Sire selection. To estimate SNP effects, the requirement is a reference population, also called training data is used (Meuwissen 2007). The reference population is comprised of at least 1000 individuals with genotype as well as phenotype information from their own performance records (refer to Fig. 6.1), deregressed proofs (VanRaden et al. 2009), and daughter-yield deviations. Genotypic and phenotypic information are combined to estimate the effects of each SNP. Genome estimated breeding values for young selection candidates are estimated using the estimates from the reference population. Several approaches have been discovered to decide which animals to include in the reference population. According to De Roos et al. (2007), VanRaden et al. (2009) proven (PT) bulls with national breeding values can be used as a reference group. Furthermore, the close relationships between young selection candidates and animals in the reference population render the Breeding Values more reliable (Habier et al. 2007).

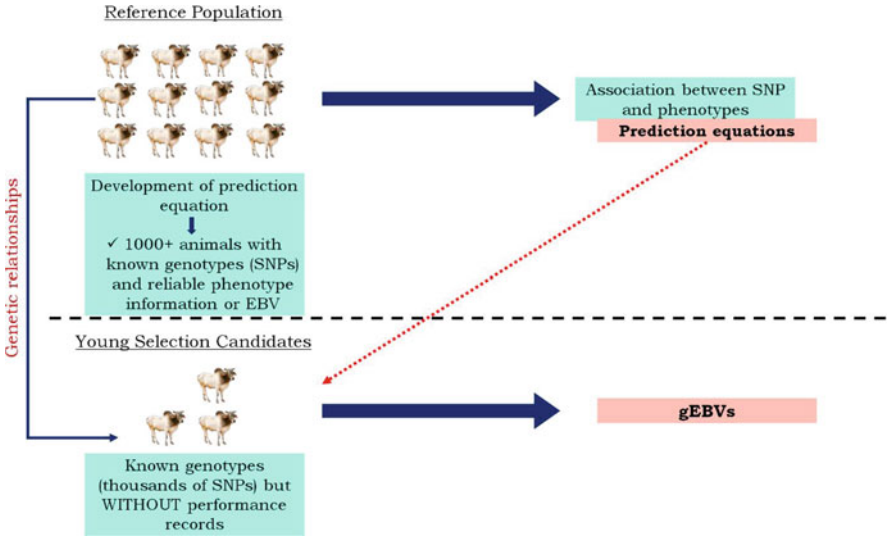

Fig. 6.1 Schematic diagram representing the overall process of prediction of GEBVs for genomic selection programs

Theoretically, to allow reliable prediction, an optimal reference population should comprise the whole range of genotypes and phenotypes. However, practically this is not possible in a real biological population, so the reference population is designed such as it reflects the whole range of phenotypes and genotypes as close as possible. A reference population composed of animals sharing a recent pedigree with the young selection candidates yields better prediction accuracy (Habier et al. 2007; De Roos et al. 2008) as LD persists across families or breeding lines used for developing prediction equation. This requires a high density of markers across the genome with high LD between markers to accurately predict the GEBVs, but the constraint lies in the cost of genotyping to generate millions of SNP genotypes in dairy Sire evaluation.

6.2.2 *Estimation of GEBVs*

Current progress in next-generation sequencing technologies has successfully created a vast pool of SNP genotype information, which along with phenotypic data and pedigree records form the simplest model to predict GEBVs (Fig. 6.1). The incorporation of pedigree records helps in deriving an additive relationship matrix (Calus and Veerkamp 2007). However, in the absence of reliable pedigree information, the additive relationship matrix is constructed directly from the genotype data (Fernando 1998). Pedigree information combined with genotypes of individuals helps in outlining possible errors in pedigree and genotype information generated. Marker haplotypes can also be derived by combining these two sources of information (Meuwissen and Goddard 2001).

Prediction of SNP effects from the reference population can be made by the most straightforward model (Calus 2010):

$$Y_i = \mu + \text{animal}_i + \Sigma\left(\text{SNP}_{ijk}\right) + e_i$$

Where,

Y_i = trait of interest or national EBV or daughter-yield deviation or deregressed proofs or average daughter performance of animal_i

μ = fixed effects or simply the mean

animal_i = polygenic random effect

Σ (SNP_{ijk}) = effects SNP, summed across allele_k ($k = 1, 2$) at loci_j for animal_i

Then, the GEBVs of animal_i can be computed in the test population: $\text{GEBV}_i = \text{animal}_i + \Sigma$ (SNP_{ijk})

Estimation of the effect of each SNP allele and fitting the allele effects ignore the recombination that often breaks the linkage between QTL and marker alleles. Incorporation of marker haplotypes that are assumed to carry the identical QTL alleles can be used in the model instead and may solve the problem to a certain extent (Meuwissen et al. 2001; Villumsen and Janss 2008). The marker haplotypes are

considered identical-by-descent (IBD) that have the identical marker alleles inherited from a common ancestor. Haplotypes become equivalent when IBD between them approaches unity, thus, it reduces the number of effects and increases the estimation power (Yu et al. 2005; Calus 2010). Quantitative trait loci (QTL) distributions suggest that there are many loci with minor effects while very few loci exert significant effects (Hayes and Goddard 2001). Haplotype or SNP effects estimated in the reference population should ideally resemble the QTL effects' underlying distribution. However, the estimated marker effects often don't capture all of the QTL effects. Additionally, smaller QTLs are not captured in QTL mapping studies and several QTLs and many SNPs linked to QTLs constitute an overall underlying distribution of the quantitative trait. Hence, care should be taken using the prior information of underlying QTL distribution to estimate the marker effects.

6.2.3 Various Methods of Solving the Model

The simple assumption behind the incorporation of SNPs or SNP haplotypes in the genetic evaluation model describing underlying QTL distribution has been explained in the previous section.

Several new methods have been developed in the past two decades for the prediction of GEBVs using genomic information. The very first method developed was a multistep model in which direct genomic values (DGV) were computed using distinct reference and validation populations. These direct genomic values could then be used in combination with conventional EBVs or simply as correlated traits (Kachman et al. 2013). This approach was used for the genomic selection of dairy cattle in the USA. It has also been applied to the genomic evaluation of the beef cattle population (Saatchi et al. 2011; Snelling et al. 2011). However, this method includes only genotyped animals in the model, and it is unmanageable to genotype all the animals.

Misztal et al. (2009) proposed the single-step method of GBLUP (ssGBLUP) that combines information on phenotypes, pedigree, and genotypes in a single evaluation. The ssGBLUP can be implemented in maternal effect and multi-trait models. It ensures that genotypic, phenotypic, and pedigree information are not double-counted, and each information source is properly weighted in the model. The ssGBLUP can also be used with a fewer number (sample) of genotyped animals. However, the multistep method lacks all this flexibility and has lesser accuracy and more genomic inflation. Aguilar et al. (Aguilar et al. 2010), Christensen and Lund (Christensen and Lund 2010) used ssGBLUP for the first time in the dairy population having more than 6000 genotyped animals.

The major challenge faced in solving the model is to accurately estimate the SNP effects from phenotypes of a few thousand animals (n) and over ten thousand to millions of SNPs (p). This results in the famous "small (n) big (p) problem" or overparameterized model; this problem is where the numbers of parameters are more than what can be estimated from the data; it can be overcome by implementing ridge

regression BLUP (RR-BLUP) or SNPBLUP. This model assumes a normal distribution of SNP effects and constant variance that implies the same proportion of variance is expressed by many if not all the SNPs, which is practically unreal. Another challenge is to eliminate the loci having very small or nearly zero effect on the distribution of QTL effects and to determine the actual number of loci whose effects need to be estimated. Weller et al. (2012) proposed the use of a subset of SNPs with large estimated effects that are incorporated in an animal model and the animal effects represent the remaining polygenic effects after the SNPs. Selecting tag-SNPs or a subset of SNPs, which explain much of the genomic variation in the complete set of SNPs genotyped, can solve the "small n big p" problem. Selection of the subset of SNPs can be achieved through various methods such as the association of SNPs with a trait of interest, feature selection using machine learning procedures (Long et al. 2007), forward stepwise regression (Habier et al. 2007), and principal component analysis (Solberg 2008).

Another important approach is to determine whether SNPs are truly associated with a QTL exerting some effect on phenotype and implement the same for the prediction of GEBV at once. Meuwissen et al. (2001) successfully implemented this approach in Bayesian models (BayesA, B, C, and π).

BayesA model assumes that all SNPs have a certain effect on the trait of interest and very few SNPs have large effects while many have small effects. Therefore, variances assumed for each SNP differ from each other. However, Bayesian methods are nonlinear and are often affected by shrinkage which makes small effects smaller and large effects even larger.

BayesB model uses a Metropolis-Hastings algorithm for every iteration to determine whether a locus has any effect on the trait or not. This method assumes that a proportion (π) of the SNPs has zero effect while 1 - π SNPs have a nonzero effect. Many loci are rendered with zero variance because of this assumption. BayesB becomes BayesA when π is zero. However, this model relies mostly on the considered prior information to estimate the SNP effects, which creates a major concern (Gianola et al. 2006).

BayesC model: As an alternative to BayesB, a mixture of SNP-BLUP and BayesB methods can be done. It assumes a constant variance for a fraction of SNPs affecting the trait as in SNP-BLUP and a fraction (π) of SNPs have zero effect as in BayesB (Meuwissen and Goddard 2004; Calus et al. 2008; Janss et al. 2008). This method is called BayesC where Gibbs sampling is applied and the Metropolis-Hastings step is avoided. QTLs with the distribution of large effects pick up the same SNPs as the BayesB model, while the SNPs with smaller effects still explain small and equal the variance as in the SNP-BLUP model (Calus et al. 2008).

BayesCπ is a modified BayesC with an additional step to estimate the proportion of SNPs not having any effect on the trait.

Other than SNP effects, the genomic relationship matrix (GRM) instead can be used in the model that assumes each of the SNP contributes equally to the additive genetic variance. Such model is known as GBLUP (Meuwissen et al. 2001). A GRM constructed over all the markers can be used in a regular mixed model instead of an additive relationship matrix (Goddard 2009). VanRaden (2008) proposed that the

contribution of each SNP to the GRM could be weighted to account for differences in variances. This method is like SNP-BLUP, but Genomic Breeding Values are estimated instead of SNP effects. The effectiveness of GBLUP depends on the closeness of GRM to the realized genetic relationships. Appropriate quality control of genomic data before constructing the GRM can avoid much of biases in any estimate. Variance not represented by the markers can be explained by including an extra polygenic effect. All these models decrease the dimensionality of the large SNP dataset.

In a decade, several alternative methods such as nonparametric kernel (Gianola et al. 2006; Gianola and Van Kaam 2008), principal component analysis, PCA (Solberg 2008), partial least squares (PLS) regression (Raadsma et al. 2008; Solberg 2008), genetic algorithms, and Bayesian-LASSO (De Los et al. 2009) have been proposed. Nonparametric kernel methods are as effective as Bayesian methods when only an additive model is assumed (Bennewitz and Meuwissen 2008), but are more effective than the Bayesian models when nonadditive (dominance and epistatic effects) effects are also considered in the model. PCA and PLS both help to reduce the dimensionality of the genotype matrix by identifying a set of markers explaining the maximum proportion of variance from the available genotypes. However, PLS depends on the trait for the same, while PCA acts independently. Apart from additive effects, dominance and epistatic (nonadditive) effects can also be included in the model, as suggested by Gianola et al. (2006) and Gianola and van Kaam (2008).

Recently, a large number of machine learning approaches such as Artificial Neural Networks (ANN), Decision Trees (DT), and Deep Learning (DL) methods have been developed for solving classification, dimension reduction, and prediction problems (Nayeri et al. 2019). ANNs consist of one or more layers of computational units called neurons mimicking the biological neural network. These neurons receive information to learn and predict based on the learned information. Deep Learning is a subtype of ANN-based method whose use is recently being prioritized for computationally cumbersome classification problems. In Decision Trees, nodes and edges are used to represent a problem. Among all the machine learning methods, DTs are the simplest methods to interpret models for prediction problems. There are several other machine learning methods, such as bagging or bootstrap aggregation (BA) (Breiman 1996), which combines multiple machine learning algorithms to make more accurate predictions, and eXtreme gradient boosting (XgBoost) (Chen and Guestrin 2016), that is focused chiefly on model performance and computational speed, and gradient boosting machine (GBM) that converts weak prediction models into stronger ones.

There exists a difference between Genome-based predicted estimated breeding values (GEBVs) and phenotype or conventional EBV predicted that the former (GEBV) uses pedigree and marker information while EBVs are predicted from progeny information. In genomic evaluations, biases may exist for the estimated effects, especially (i) when QTLs are estimated as fixed effects and (ii) when there is insufficient LD between the marker and QTL, as the latter may lead to underestimation of the effects. Practically, this arises as all the SNPs present on an SNP chip are not utilized for prediction purposes. VanRaden et al. (2013) suggested that about

40% of the total SNPs present on a chip are useful for prediction and explain the reasons for a smaller number of usable SNPs as their low call rates, lower minor allele frequency (MAF), absence of sufficient polymorphism, and redundancy of the markers.

An important concern in genomic evaluation is to decide the nature of the marker effect—whether it should be included in the model as a fixed or random effect. SNPs are often considered a fixed effect in various studies (Baruch and Weller 2008), especially when the sample genotyped is very small. However, recent studies show that when marker effects are treated as random, it eliminates the problem of over-parameterization (when the number of genotypes outnumbers the records/sample) in the prediction. The random markers effect also takes into account the covariances among the marker effects and minimizes the 'noise' or null effect of markers on the trait of interest (Weller 2016).

6.2.4 Accuracies of Genomic Predictions

The accuracy of the GEBV prediction depends upon several factors, and these should be taken into consideration while developing the prediction model and designing the genomic evaluation programs. Of course, there are some ways to improve the accuracy of a genomic prediction.

Independent of the prediction approach, one such major factor is the linkage disequilibrium, LD, between the marker and the QTL and its consistency in the reference as well as the test population (Goddard 2009). The LD between the marker and QTL mostly depends on the marker density. If the LD is incomplete or broken in further generations, reestimation of marker effects is required; but some of the QTL variances still can't be captured and are left out even after a reestimation attempt is made; thus, incomplete LD leads to lower accuracy. In addition to the above, the accuracy of the marker effects estimated depends on the number of phenotypic records included in the reference population (Hayes et al. 2009; VanRaden et al. 2009), selection of animals for the reference population (Habier et al. 2007), and the heritability of the target trait (Calus et al. 2008; Hayes et al. 2009).

It is imperative that the accuracy can be improved by including more animals in the reference population (Meuwissen et al. 2001). The nature of sampling, by selecting the individuals within a family for both reference and test population in such a way that the same genetic relationship exists between the individuals of two populations, can potentially improve the accuracy of genomic prediction (Legarra et al. 2008). The accuracy of GEBVs can also be increased by increasing the marker density (Meuwissen et al. 2001; Calus et al. 2008; Solberg et al. 2008) and by increasing the heritability of the trait (Calus et al. 2008; Solberg 2008).

In dairy-Sire evaluation with the real dataset, when unequal variance for each SNP is assumed, the GEBV predicted is slightly more reliable than the equal variance assumption (Hayes et al. 2009; VanRaden et al. 2009). BayesB gives more reliable estimates (VanRaden et al. 2009). Single-step GBLUP (ssGBLUP)

yields the least biased GEBVs, more consistent solutions, and better accuracy as it integrates phenotype, genotype, and pedigree data. However, its limitation lies in its assumption of equal variances for each of the SNPs.

An additional gain in accuracy can be obtained by optimizing weights for SNPs in SNP-weighted ssGBLUP (WssGBLUP). National EBVs or Degressed proofs can be used instead of raw phenotypes to increase the accuracy as the heritability is effectively increased per animal due to the weighting of the offspring contribution to bull's EBV (Fikse and Banos 2001). Accuracies can also be improved by implementing cross-validation methodologies in Bayesian models (BayesB) and genomic relationship matrices (G-BLUP) (Habier et al. 2010). Ferdosi et al. (2019) proposed a new method that expedites the calculation of accuracy for new individuals with genotypes, with or without phenotypes, without recalculating the inverse matrix for the whole population by using the previously calculated matrices. Some other studies describe that the accuracy of predictions depends on the number of independent chromosome segments (genomic structure) associated with the economic trait (Daetwyler et al. 2010).

6.3 Data Repositories Available in India and their Effective Use

The advancement in the use of ICT, mobile phones, and tablets renders novel opportunities to initiate large-scale phenotype recordings across the country. The data recording system should be simple, efficient, cost-effective, non-evasive, and easily accessible. The information on animal performance, welfare, feed efficiency, and product quality constitutes a large dataset; hence, a sturdy and effective collection and storage database should be built. A suitable technique should render routine data collection, real-time transmission to the mainframe database, and robust analytics providing feedback and access to users. The storage database should have user-friendly interface and should support the storage of both phenotypic and genotypic information to support genomic selection and optimize livestock production systems. Information collected across the country provides a range of production environments that facilitate the selection for economic traits suitable for the specific agro-climatic zones. This will increase the incorporation of such data in genetic improvement programs. Management decisions for farmers can also be formulated based on this information that would encourage them to actively participate in the data recording systems.

However, in the Indian scenario, a reliable internet connection poses a greater challenge for on-farm and field level recordings at smallholders' farmer's herd. The development of offline technologies for data capture would overcome such challenges. Emerging cheap and innovative technologies like the mid-infrared spectrum can be used for indirect prediction of economically important traits in dairy cattle, and when calibrated and adapted properly, will benefit hugely in the smallholder

systems (Burrow et al. 2021). However, a large-scale interdisciplinary collaboration of animal scientists with software developers and engineers is needed to develop such an efficient data recording system across the country.

Non-Governmental Organizations such as the Bharatiya Agro Industries Foundation (BAIF) initiated a dairy recording program of crossbred cows as well as crossbred and pure-exotic bulls in the artificial insemination (AI) stud across six states (Bihar, Jharkhand, Maharashtra, Odisha, and Punjab Uttar Pradesh) in India, to generate estimates of breed composition and GEBV (Al Kalaldeh et al. 2021). It involved ~170 AI technicians with "data loggers"and mobile phones installed with multicomponent software. Ducrocq et al. (2018) reviewed some of the milk recording systems in India by the National Dairy Development Board (NDDB), working in collaboration with several developmental agencies and BAIF. However, they reported the recording to be immature with a high percentage of incomplete records and the use of poor quality bulls under natural mating system rendering the problem of unknown sire, animal identification errors, and high human error.

A large number of information on daughter performances of test bulls under different progeny testing (PT) programs are available and sires are being evaluated regularly based on the information available from various coordinating farm units and from participating farmers' herd at the field level. If all the daughters of test bulls could be genotyped, both the genotype and phenotype information from progeny testing programs of different breeds and species can be included to form a database. Under Network Project on Buffalo Improvement for Murrah buffaloes, authors evaluated the daughter records available at both farm and field levels. A total of 15 buffalo test bulls were evaluated based on an average of 56 daughter records per sire. Authors tested several BLUP Animal Models, such as, 1. An animal model with age at first calving (AFC) and season of calving as a fixed effect, 2. An animal model with the season of calving as a fixed effect, 3. An animal model with period and season of calving as fixed effects, and 4. An animal model with AFC, season, and period of calving as fixed effects. It was observed that the model with AFC, season, and period of calving as fixed effect performed best with the lowest log Likelihood and Akaike Information Criterion (AIC) value. In these models, if genotype information of all the daughters and test bulls had been included, this particular set under the PT program would have turned into a reference population for genomic selection. Along with that, if the genotype information could be generated for all the breeds under different PT programs, a multi-breed reference population can also be formed. Although it would be computationally cumbersome to select animals based on multi-breed reference population, it will capture the wide geographical variation within a species in India. Yet, to implement genomic selection in Indian conditions, small herd sizes have always been a big hindrance. However, Powell et al. (2021) demonstrated that genomic evaluation programs would be feasible even with dairy cattle populations with weak genetic connectedness, small herd sizes, and low heritability of economic traits. Accuracies of EBV estimation can be improved by modeling small herds as random effects in genomic evaluations (Visscher and Goddard 1993; Schaeffer 2009; and Wiggans et al. 2017). Hence, there is great scope for genomic improvement of smallholder dairy herds compared to pedigree-

based evaluations as there is an increase in genetic connectedness between tiny herds due to shared haplotypes rather than shared relatives (Powell et al. 2021). The opportunity to use genomic information along with vast pedigree records already available and new records generated via data recording systems with integrated ICTs offers a possibility to implement genomic selection to increase the rate of genetic gain. This would also benefit in the characterization of indigenous and crossbred animals for conservation purposes and formulating of crossbreeding and within-breed selection programs to improve economically important traits. Genomic selection is still a naive yet emerging field in India, which can transform the future of genomic selection in India through the use of many phenotype records available in India (Vohra 2018).

6.4 Challenges and Future Prospects

The accuracy of GEBVs predicted is high; to obtain maximum genetic gain, these GEBVs should be incorporated into National Evaluation Programs without further delay (Harris et al. 2008). This depends on the adaptability of National Evaluation Services and dairy industries to include genomic information for the prediction of Breeding Value and selection of Sires with high genetic merit.

When we consider the dairy industry, then the industry-wide scale of combining pedigree, phenotype, and genomic information to predict GEBVs poses a considerable challenge. One of the major challenges lies in the limited number of animals genotyped among the population. In dairy-advanced countries, genomic selection (GS) has resulted in an increased rate of genetic gains in dairy cattle and the production of genomically proven bulls. The credit for their huge success (implementation of GS) can be attributed to the fact that a well-established conventional genetic evaluation system was already in place long before the introduction of GS. On the other hand, developing countries are lagging in implementing GS, maybe due to a lack of proper breeding infrastructure, pedigree, systematic phenotypic recording, and the absence of computational and analytical tools, which are fundamental for the success of any conventional genetic evaluation programs. Ribaut et al. (2010) indicated the upsurge in information and communication technology has created opportunities to counter some of the challenges specific to developing countries by establishing global virtual platforms. The rapid developments in high-throughput next-generation sequencing technologies for SNP genotyping have reduced the genotyping costs. There are few consortiums working solely for genomic evaluation in dairy cattle in developing countries. Cost-effective genotyping services provided by several companies may boost the implementation of genomic evaluations, yet it is necessary to establish consortiums for genomic evaluation programs on a large scale in developing countries of South-Asian and African subcontinents (Mrode et al. 2019).

Genomic selection programs should also include tools for parentage verification, determination of breed composition, breed characterization, and computation of

genomic inbreeding coefficients for sustainable utilization of genomic information. Such an approach enhances the cost-effectiveness of genotyping and provides maximum benefit. Higher returns from the genomic selection are expected by implementing artificial reproduction technologies, such as AI and IVF in dairy cattle, which would aid in the dissemination of superior germplasm, gender-biased semen from genomically proven bulls. This shall augment the productivity and profitability of farmers having small animal holdings. Farmers and breeder associations should come forward and be encouraged to participate in nationwide Sire evaluation programs. Hand-holding of developing countries by dairy-developed nations for the implementation of genomics-based Sire selection should be promoted. Use of a Sire produced and evaluated in one country can be easily used in female population across the nations as nowadays transportation of bull's germplasm is few hours distance. The role of Governments to formulate the breeding policy favoring the GS and regulatory frameworks for initiating long-term genomic breeding programs all shall play a decisive role in the overall success of the GS programs. Further, it is proposed that the limited genotype information available in each country could be pooled together based on the geographical distribution of the breed/transboundary breeds, which would possibly increase the accuracy at the initial stages (Mrode et al. 2018). Simultaneously, the issues related to data ownership and accessibility from pooling the genomic data should also be addressed. Thus, governing bodies formed by joint efforts of various countries should manage data sharing and float the confidentiality agreement.

There is growing concern about pedigree-based genetic evaluation analyses, as being less efficient and accurate than genomic methods. However, linking pedigree data to information obtained through genomics appears to be advisable, since it may significantly improve the accuracy of predicting traits and indices in dairy breeds (Aguilar et al. 2010).

6.5 Conclusion

Genomic selection is the ultimate approach that exploits molecular markers, SNPs, to evaluate dairy animals and design novel breeding programs. Genomic breeding values for young selection candidates can be predicted from genotype information and a prediction model can be developed and tested in the reference populations. The gain inaccuracies due to genomic evaluations is up to 0.31 higher than that of pedigree-based indices. Several methods have been reported to estimate marker effects and predict the GEBV. However, the prediction accuracy depends upon LD between the marker and QTL, size of the reference population, the genetic relationship between the reference and test population, the heritability of the trait, the prediction model used, and the number of animals genotyped. There has been an increase in reliabilities of GEBV over conventional EBV (2–20%) for young bulls with no daughter records. It has been seen that the higher genetic gain in genomic selection programs is consequent to reduced generation interval when compared to

conventional progeny testing programs. Although genomic selection has started yielding fruitful results in the past decade, challenges and opportunities both remain in the implementation of GS programs. In times to come, it is envisaged that with the advent of biotechniques for direct gene manipulation through CRISPER/Cas9 technology and others, the possibility also exists to change the genotypes of embryos in farm animals. GS and other technologies have a bearing on future Sire evaluation and selection in a breeding program. The story of 'omics' has just begun.

References

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ (2010) Hot topic: a unified approach to utilize phenotypic full pedigree and genomic information for genetic evaluation of Holstein final score. J Dairy Sci 93(2):743–752

Al Kalaldeh M, Swaminathan M, Gaundare Y, Joshi S, Aliloo H, Strucken EM, Gibson JP (2021) Genomic evaluation of milk yield in a smallholder crossbred dairy production system in India. Genet Sel Evol 53(1):1–14

Baruch E, Weller JI (2008) Incorporation of discrete genotype effects for multiple genes into animal model evaluations when only a small fraction of the population has been genotyped. J Dairy Sci 91(11):4365–4371

Bennewitz J and Meuwissen TH (2008) Genomic breeding value estimation using kernel regression and additive models. In: 12th quantitative trait locus and marker assisted selection workshop Uppsala Sweden 34

Breiman L (1996) Bagging predictors. Mach Learn 24:123–140

Burrow HM, Mrode R, Mwai AO, Coffey MP, Hayes BJ (2021) Challenges and opportunities in applying genomic selection to ruminants owned by smallholder farmers. Agriculture 11(11): 1172

Calus MP (2010) Genomic breeding value prediction: methods and procedures. Animal 4(2): 157–164

Calus MPL, Veerkamp RF (2007) Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. J Anim Breed Genet 124(6):362–368

Calus MPL, Meuwissen THE, De Roos APW, Veerkamp RF (2008) Accuracy of genomic selection using different methods to define haplotypes. Genetics 178(1):553–561

Chen T, Guestrin C (2016) Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 785–794

Christensen OF, Lund MS (2010) Genomic prediction when some animals are not genotyped. Genet Sel Evol 42(1):1–8

Crick F, Watson J (1953) A structure for deoxyribose nucleic acid. Nature 171(737–738):3

Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The impact of genetic architecture on genome-wide evaluation methods. Genetics 185(3):1021–1031

De Los CG, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, Cotes JM (2009) Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics 182(1):375–385

De Roos APW, Schrooten C, Mullaart E, Calus MPL, Veerkamp RF (2007) Breeding value estimation for fat percentage using dense markers on Bos taurus autosome 14. J Dairy Sci 90(10):4821–4829

De Roos APW, Hayes BJ, Spelman RJ, Goddard ME (2008) Linkage disequilibrium and persistence of phase in Holstein–Friesian Jersey and Angus cattle. Genetics 179(3):1503–1512

Ducrocq V, Laloe D, Swaminathan M, Rognon X, Tixier-Boichard M, Zerjal T (2018) Genomics for ruminants in developing countries: from principles to practice. Front Genet 9:251

Ferdosi MH, Connors NK, Tier B (2019) An efficient method to calculate genomic prediction accuracy for new individuals. Front Genet 10:596

Fernando RL (1998) Genetic evaluation and selection using genotypic phenotypic and pedigree information. In: Proceedings of the 6th world congress on genetics applied to livestock production, vol 26, Armidale, pp 329–336

Fikse WF, Banos G (2001) Weighting factors of sire daughter information in international genetic evaluations. J Dairy Sci 84(7):1759–1767

Fisher RA (1918) The correlation between relatives on the supposition of mendelian inheritance. Trans Royal Soc Edinburgh 52:399–433

García-Ruiz A, Cole JB, VanRaden PM, Wiggans GR, Ruiz-López FJ, Van Tassell CP (2016) Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. Proc Natl Acad Sci 113(28):E3995–E4004

Gianola D, Van Kaam JB (2008) Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. Genetics 178(4):2289–2303

Gianola D, Fernando RL, Stella A (2006) Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics 173(3):1761–1776

Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136(2):245–257

Habier D, Fernando RL, Dekkers JC (2007) The impact of genetic relationship information on genome-assisted breeding values. Genetics 177(4):2389–2397

Habier D, Tetens J, Seefried FR, Lichtner P, Thaller G (2010) The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet Sel Evol 42(1):1–12

Harris BL, Johnson DL, Spelman RJ and Sattler JD (2008) Genomic selection in New Zealand and the implications for national genetic evaluation. In: Proc Interbull Meeting Niagara Falls Canada

Hayes BEN, Goddard ME (2001) The distribution of the effects of genes affecting quantitative traits in livestock. Genet Sel Evol 33(3):1–21

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: Genomic selection in dairy cattle: Progress and challenges. J Dairy Sci 92(2):433–443

Janss L, GregersenV, Bendixen C and Lund M (2008) Validation of genomic predictions in pigs using medium-dense marker coverage. In: Book of abstracts of the 59th annual meeting of the EAAP Vilnius Lithuania

Kachman SD, Spangler ML, Bennett GL, Hanford KJ, Kuehn LA, Snelling WM, Pollak EJ (2013) Comparison of molecular breeding values based on within-and across-breed training in beef cattle. Genet Sel Evol 45(1):1–9

Legarra A, Robert-Granié C, Manfredi E, Elsen JM (2008) Performance of genomic selection in mice. Genetics 180(1):611–618

Long N, Gianola D, Rosa GJ, Weigel KA, Avendaño S (2007) Machine learning classification procedure for selecting SNPs in genomic selection: application to early mortality in broilers. J Anim Breed Genet 124(6):377–389

Meuwissen T (2007) Genomic selection: marker assisted selection on a genome wide scale. J Anim Breed Genet 124:321–322

Meuwissen TH, Goddard ME (2001) Prediction of identity by descent probabilities from marker-haplotypes. Genet Sel Evol 33(6):1–30

Meuwissen TH, Goddard ME (2004) Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data*Genetics selection*. Evolution 36(3):261–279

Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4):1819–1829

Misztal I, Legarra A, Aguilar I (2009) Computing procedures for genetic evaluation including phenotypic full pedigree and genomic information. J Dairy Sci 92(9):4648–4655

Mrode R, Coffey M, Ojango J, Mujibi D, Okeyo M, Strucken EM, Aliloo H (2018) The impact of modelling and pooled data on the accuracy of genomic prediction in small holder dairy data. In: Proceedings of the World Congress on Genetics Applied to Livestock Production. Auckland

Mrode R, Ojango JM, Okeyo AM, Mwacharo JM (2019) Genomic selection and use of molecular tools in breeding programs for indigenous and crossbred cattle in developing countries: current status and future prospects. Front Genet 9:694

Nayeri S, Sargolzaei M, Tulpan D (2019) A review of traditional and machine learning methods applied to animal breeding. Anim Health Res Rev 20(1):31–46

Powell O, Mrode R, Gaynor RC, Johnsson M, Gorjanc G, Hickey JM (2021) Genomic evaluations using data recorded on smallholder dairy farms in low-to middle-income countries. JDS Commun 2(6):366–370

Raadsma HW, Moser G, Crump RE, Khatkar MS, Zenger KR, Cavanagh JAL, Tier B (2008) Predicting genetic merit for mastitis and fertility in dairy cattle using genome wide selection and high density SNP screens. In: Animal genomics for animal health, vol 132, pp 219–223

Ribaut JM, De Vicente MC, Delannay X (2010) Molecular breeding in developing countries: challenges and perspectives. Curr Opin Plant Biol 13(2):213–218

Saatchi M, McClure MC, McKay SD, Rolf MM, Kim J, Decker JE, Taylor JF (2011) Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. Genet Sel Evol 43(1):1–16

Sattler CG (2013) Progeny testing and genomics: where are we and where are we going? In: Proceedings Dairy Cattle Reproduction, Council Annual Meeting Indianapolis, pp 98–103

Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet 123(4):218–223

Schaeffer LR (2009) Contemporary groups are always random. http://www.aps.uoguelph.ca/~lrs/LRSsite/ranfix.pdf

Snelling WM, Allan MF, Keele JW, Kuehn LA, Thallman RM, Bennett GL, Rolfe KM (2011) Partial-genome evaluation of postweaning feed intake and efficiency of crossbred beef cattle. J Anim Sci 89(6):1731–1741

Solberg TR (2008) Methods for prediction of genome-wide breeding values using dense marker genotyping, vol 41. Norwegian University of Life Sciences

Solberg TR, Sonesson AK, Woolliams JA, Meuwissen THE (2008) Genomic selection using different marker types and densities. J Anim Sci 86(10):2447–2454

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91(11): 4414–4423

VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS (2009) Invited review: Reliability of genomic predictions for North American Holstein bulls. J Dairy Sci 92(1):16–24

VanRaden PM, Null DJ, Sargolzaei M, Wiggans GR, Tooker ME, Cole JB, Doak GA (2013) Genomic imputation and evaluation using high-density Holstein genotypes. J Dairy Sci 96(1): 668–678

Villumsen TM and Janss L (2008) Genomic selection focusing on haplotype length and prior settings. In: 12th quantitative trait locus and marker assisted selection workshop Uppsala, Sweden 41

Visscher PM, Goddard ME (1993) Fixed and random contemporary groups. J Dairy Sci 76(5): 1444–1454

Vohra V (2018) Genomic selection and its significance in Indian dairying. Indian J Dairy Sci 71(6): 539–545

Weller JI (2016) Genomic selection in animals. John Wiley and Sons Incorporated

Weller JI, Glick G, Shirak A, Ezra E, Zeron Y, Ron M (2012) Predictive ability of selected subsets of single nucleotide polymorphisms (SNPs): for moderately sized dairy cattle populations. Interbull Bulletin 8(2):208–216

Wiggans GR, Cole JB, Hubbard SM, Sonstegard TS (2017) Genomic selection in dairy cattle: the USDA experience. Annu Rev Anim Biosci 5:309–327

Yu K, Xu J, Rao DC, Province M (2005) Using tree-based recursive partitioning methods to group haplotypes for increased power in association studies. Ann Hum Genet 69(5):577–589

Zhang Z, Zhang Q, Ding X (2011) Advances in genomic selection in domestic animals. Chin Sci Bull 56(25):2655–2663

Chapter 7
NGS-Based Biomarkers in Livestock

Chandra Sekhar Mukhopadhyay and Bhawanpreet Kaur

Abstract Introduction of high-throughput molecular technologies and associated bioinformatic tools that potentially made scientists to acquire, process, and analyze large amounts of omics data has evolving significantly. NGS-based biomarkers in livestock allow detailed studies at the genome (genomics), transcriptome (transcriptomics), metabolome (metabolome), proteome (proteomics), MiRNA, lncRNA, molecular markers, genomic selection (GS), and role of genome-wide association studies (GWAS) in cattle, buffalo, chicken, horse, pig, and goat. Using various next-generation sequencing (NGS) platforms, animal genomes have been sequenced. The quick and accurate collection of genome sequences has enabled the discovery of causative common and unusual changes across the genome.

The goal of this chapter is to provide a current and prospective outlook on the benefits and challenges associated with the value of NGS-based biomarkers in animal genetics, breeding with a particular focus on disease, molecular analysis, vaccine production, epidemiological studies, infectious disease diagnosis, and management.

Keywords Biomarkers · Sequencing · Molecular · Livestock · Genetics · Breed

7.1 Introduction

Biomarkers are defined as biological markers, an important tool used to determine the functioning of the body via physical signs or measures. In other words, biomarkers refer to molecules that are common but not always part of the biological mechanism. Biomarkers can be any type of molecules, such as RNA, proteins, or metabolites, or they might be a profile of many molecules. The biomarkers aim to capture the overall event happening in a particular cell or an organism in a given

C. S. Mukhopadhyay (✉) · B. Kaur
Department of Bioinformatics, College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India
e-mail: csmukhopadhyay@gadvasu.in

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_7

moment. These are the strong detector, provide an early warning sign in the body or even an indicator of normal biological, pathogenic processes, or pharmacological responses to a therapeutic intervention (Naylor 2003). In practice, biological markers include tools and technologies that can help in recognizing the cause, determining, diagnosis, and outcome of treatment of diseases. To study human diseases, livestock diseases, different types of biomarkers have been used by generations of scientists, epidemiologists, and physicians.

Biomarkers are derived from a better understanding of the biological mechanism behind a commercially valuable feature. The level of the biomarker is frequently linked to quantitative features of the characteristic (Te Pas et al. 2010; Wilson and De Vega 2005). The level of specific antibodies, for example, could be a good biomarker (indicator) for disease protection induced by vaccination. There is immense applications of biomarkers in the immunological diseases, genetic disorders, infections, diagnosis and management of cardiovascular disease, cancers (Perera and Weinstein 2000). Molecular biomarkers provide a dynamic and powerful approach to understanding the spectrum of disease with obvious applications in analytic epidemiology, clinical trials, disease prevention, diagnosis, and disease management. Because of the tremendous advancements in molecular biology and laboratory technology, the use of technically advanced biomarkers will soon become even more practical.

Biomarkers are split into two major types: biomarkers of exposure, which are used in risk prediction, and biomarkers of disease, which are used in disease screening, diagnosis, and monitoring.

Many biomarker studies fall short of their full potential due to a failure to follow the same standards that would apply to the use of nonbiological variables. Any biomarker development should take place before or concurrently with the standard design of any epidemiological study or clinical trial. Pilot studies must be done to establish the accuracy, reliability, interpretability, and practicality of the laboratory component. The researcher must create "normal" distributions for key variables like age and gender. The researcher will also wish to determine the biomarker's intraindividual variation, tissue localization, and persistence. Biomarkers are well established in risk prediction, screening, and diagnostic procedures, and they provide distinct and evident benefits. Many neurological disorders are classified using either established clinical criteria or histology diagnoses.

Biomarkers offer the ability to identify neurological disease at an early stage, to give a mechanism for homogeneous disease classification, and to expand our understanding of disease causation. These benefits apply to all types of clinical research, from clinical trials to epidemiological observational studies. Biomarkers used in risk prediction, screening, and diagnostic tests are well established, and they offer distinct and obvious advantages. The classification of many neurological diseases is based on either standardized clinical criteria or histological diagnoses. Biomarkers also have the potential to identify neurological disease at an early stage, to provide a method for homogeneous classification of disease, and to extend our knowledge base concerning the underlying disease pathogenesis. These advantages have direct application to all types of clinical investigation, from clinical trials to

Table 7.1 Types of biomarkers and their features

Types of biomarkers	Features
Molecular biomarkers	Have biophysical properties, which allow their measurements in biological samples
Radiographic biomarkers	Obtained from imaging studies
Diagnostic biomarkers	Detects the presence or identifies an individual with a subtype of the disease
Physiologic biomarkers	Measures of body processes
Monitoring biomarkers	Detect an effect of a medical product or biological agent
Prognostic biomarkers	Identify the likelihood of a clinical event, disease recurrence, or disease progression in patients
Histologic biomarkers	Reflect a biochemical or molecular alteration in cells, tissues, or fluids
Predictive biomarkers	Finding that the presence or change in a biomarker predicts whether an individual or a group of individuals will have a favorable or unfavorable reaction to a medical product or an environmental agent
Digital biomarkers	Sensors and personal devices now enable rapid and continuous assimilation of information about a person that provides insight into complex measures

observational studies in epidemiology. Biomarkers are useful for revealing connections between environmental exposures, human biology, and disease. Biomarkers can help scientists better understand basic biological processes, develop exposure science, and translate research findings into medicinal and public health applications (Table 7.1).

Key Features of Biomarkers

- For illness status, biomarkers are accurate, sensitive, and specific.
- Unaffected by other diseases biomarker.
- Quantification of the biomarker from accessible body fluids or tissues is accurate.
- Biomarker abundance is not subject to large variations in the overall population.
- Measurements that are repeatable and reliable in a variety of situations and at different times.
- The results of biomarkers are simple to understand.

7.2 Proteomic Biomarkers in Livestock

7.2.1 Salivary Biomarkers

The fluid in the oral cavity is known as saliva. Saliva is an excellent source of indicators associated with stress and systemic disease. It comprises a variety of proteins that differ between animals, indicating different diets and digestion styles. The new stress-related proteins discovered using a high-resolution quantitative proteomic method have the potential to be novel salivary biomarkers for stress

diagnosis and prevention (Escribano et al. 2019). Saliva is increasingly being used to assess biomarkers in a variety of species (Rubio et al. 2019). Antibody generation and efficacy of vaccines against anthrax, listeriosis, paratuberculosis, and footrot illnesses improved after these protein biomarkers were included in vaccine development. After a stressful event like grouping, there is a significant increase in oxidant biomarkers like AOPP and FOX as well as antioxidant biomarkers like CUPRAC, FRAS, and TEAC in the saliva of calves (Rubio et al. 2021).

7.2.2 Milk Biomarkers

Biomarkers in milk are being hailed as a promising new tool for animal health management Roncada et al. 2012. Early detection of mastitis in goats and sheep, as well as identification of pathogenic and host immune response investigations, have all benefited from biomarkers.

Overall, milk biomarkers can help with the development of new management strategies to boost animal productivity and public health by boosting early disease identification and prevention. In fresh dairy cows, milk lactose content can be used as a health indicator. Cows with greater lactose levels also have higher glucose levels. A low lactose content can indicate mastitis and metabolic diseases (TeleviCius et al. 2021).

7.2.3 Seminal Biomarkers

The most abundant metabolites in bull seminal plasma are fructose, citric acid, lactic acid, urea, and phosphoric acid. In the recent study, the metabolite profiles between high and low fertility bulls, with fructose and 2-oxoglutaric acid as potential indicators of bull fertility and acts as potential biomarker (Velho et al. 2018). The sncRNAs play critical role in spermatogenesis, sperm maturation and embryo development. Besides, these are being used as potential fertility biomarkers.

7.2.4 Urinary Biomarkers

Diarrhea, which affects fluid and electrolyte balance, is one of the leading causes of neonatal mortality in calves. Proteins involved in the renal transport of water and electrolytes were also discovered, which could aid in the early detection and prevention of illnesses (Dratwa-Chałupnik et al. 2020).

Urinary biomarkers were investigated for the diagnosis, pathogenesis, and monitoring therapy of various diseases such as kidney, liver, diabetes, neurodegenerative disorders, and prion disease (Pisitkun et al. 2006). The detection of pregnancy in the

early stages has also been possible with the help of urinary biomarkers. Urinary uromodulin possesses the biomarker potential to be used for early detection of pregnancy in cattle. The Peptide profiling in cow urine describes the molecular signature of several physiology-driven pathways as well as bioactive characteristics. The presence of urinary peptides, precursor proteins, and proteases provides a baseline of information in healthy cows that can be used in biomarker identification research for a variety of pathophysiological studies (Kumar et al. 2021).

7.2.5 Tears Biomarkers

The comparative proteomic analyses of tears from healthy humans, cows, sheep, and camels were first reported. Differential protein expression existed in the tear among species, offering useful information for further study on tear proteins and the related ocular diseases (Shamsi et al. 2011). The protein analysis of rabbit tears revealed that two defensins, NP-1 and NP-2, were significantly increased during the curative process following mechanical abrasion of the cornea, and thus may play a critical role in the cornea's protection against microbial infection during the healing phase.

7.3 MicroRNAs as Biomarkers

MiRNAs are a type of noncoding single-stranded RNA that plays a key function in the posttranscriptional regulation of gene expression. They are small (19–24 nucleotides in length) and noncoding. miRNAs are good candidates as disease biomarkers as their expression changes in various disease states such as autoimmune, neurodegenerative, and inflammatory diseases. They interact with their target genes' 3′--translated regions (UTRs) and cause posttranscriptional downregulation of those genes through target mRNA cleavage, deadenylation, and/or translation inhibition (Wu et al. 2006). They also attach to the 5′UTR of target mRNA, causing gene up- or downregulation including steroid hormone production and release (Nothnick 2012; Orom et al. 2008; Place et al. 2008).

MicroRNAs (miRNAs) are good for disease biomarkers.

1. miRNA expression profiles tend to be distinct across conditions and do not appear to be affected by age, race, or gender.
2. Most miRNA sequences are conserved across species.
3. The expression of several miRNAs appears to be tissue or biological state-dependent.
4. Manipulating various miRNAs linked with malignancies in vivo mice models.
5. Various studies show that circulating levels of miRNAs return to baseline after treatment, showing their potential utility as a therapeutic target.

Cows and pigs have been identified to have some miRNAs that could be used as biomarkers for stress management and regulation, as well as animal health control, as our understanding of miRNAs advances (Shaughnessy et al. 2020). MiRNAs may potentially be used as biomarkers for stress resilience or sensitivity (Chen et al. 2015). Mirna indicators linked to stressors can be found in biological matrices like milk, saliva, tears, cerebrospinal fluid, urine, and hair.

Researchers successfully characterized plasma miRNA populations linked with pregnancy in cattle using genome-wide analysis (GWA). Their use in reproductive biology has allowed researchers to seek pregnancy indicator molecules in dairy cattle. In the future, miRNA biomarkers are in great demand in the early detection of pregnancy (Lim et al. 2021). Several miRNAs have been associated with diseases in farm animals which affect livestock productivity and health.

In the study, Ioannidis and Donadeu (2018) found 92 miRNAs were more abundant in plasma than in cells. MiR-802, a previously unidentified bovine miRNA, was shown to be substantially enriched in the liver after expression profiling of selected plasma-enriched miRNAs across diverse bovine tissues. The miRNA database created in the study will be valuable for future research into miRNA biology in cattle and the discovery of additional possible tissue function biomarkers in livestock.

MiRNAs are a good candidate to distinguish the infected/noninfected milk status of cows. Milk from mastitis-affected cows contains different inflammatory-associated miRNAs than milk from healthy cows. The differential expression of three miRNAs (MIR29B-2, MIR146A, and MIR155) was investigated. The MIR29B-2 appears to be a promising biomarker for bovine mastitis. MIR29B-2's high specificity, sensitivity, and accuracy could be useful in predicting the state of milk (infected vs. noninfected) from animals with unclear illness status Srikok et al. 2020. MicroRNAs are becoming increasingly important in animal reproduction. MicroRNAs are expressed differently in pregnant and nonpregnant buffalo. MiRNAs are being evaluated as prospective biomarkers for a variety of human disorders, including autoimmune (Pauley et al. 2009), metabolic (Rottiers and Näär 2012), and cardiovascular diseases as well as several types of malignancies (Alvarez-Garcia and Miska 2005) due to their critical regulatory functions in many biological processes across species. The table shows the number of miRNAs in different animals. The major database available for miRNA is MiRBase release 21 (http://www.mirbase.org) (Kozomara and Griffiths-Jones 2014) and NONCODE database 2016 (www.noncode.org) (Zhao et al. 2016) (Table 7.2).

Table 7.2 Detected the number of miRNAs in farm animals

Species	miRNA	
	Precursor	Mature
Cattle	808	793
Sheep	106	153
Goat	267	436
Pig	382	411
Chicken	740	994
Mouse	1193	1915

7.4 Long Noncoding RNAs (lncRNA) as Biomarkers

lncRNAs are involved in a variety of areas of normal physiology and disease, including gene expression regulation. Noncoding RNA is classified into two groups: short (sncRNA) and long noncoding (lncRNA) (Nie et al. 2012). Noncoding RNAs longer than 200 nucleotides are classified as lncRNAs (Kozomara and Griffiths-Jones 2014). This cutoff is used to distinguish lncRNAs from short RNAs such as microRNAs (miRNAs) and transfer RNAs. Long-intergenic ncRNAs (lincRNAs), intronic lncRNAs, antisense lncRNAs, sense lncRNAs, enhancer-associated lncRNAs, and circular lncRNAs are the six different forms of lncRNAs (Laurent et al. 2015). Some lncRNAs that contain introns that can be nonpolyadenylated, or are bidirectionally transcribed have shown differences. Conservation and homology are also crucial. Only 12% of the genome can code for proteins. The ncRNA molecules are referred to as "transcription noise" because of their typically low expression levels (Ma et al. 2013). lncRNAs are significant in genomic studies, identifying them in farm animal genomes is essential for bridging the genotype-phenotype gap. Browser (http://genome.ucsc.edu/) and the NONCODE database (http://www.noncode.org/) can be used to find conserved lncRNAs across species. Broadly, the lncRNAs Can Be Divided into Five Categories:

- Sense lncRNAs
- Antisense lncRNAs
- Bidirectional lncRNAs
- Intronic lncRNAs
- Intergenic lncRNAs

7.4.1 lncRNA in Bos taurus

Huang et al. (2012) discovered 449 lncRNAs in 405 intergenic areas using public bovine-specific expressed sequence tag sequences, resulting in the first genome-wide catalog of bovine intergenic lncRNAs. Koufariotis et al. (2015) published a complete genome-wide annotation of lncRNA expression in 18 different tissues. The majority of studies of lncRNAs were related to their expression in the mammary gland (Cai et al. 2018), milk exosomes (Zeng et al. 2019) concerning dairy production. The long noncoding transcriptome of male reproduction traits was analyzed (Wang et al. 2019) (Table 7.3).

7.4.2 lncRNA in Gallus gallus

When comparing the complexity of human and chicken transcriptomes found that the chicken transcriptome is similar to the human transcriptome. The studies relate to

Table 7.3 Genome-wide scale long noncoding RNA (lncRNA) and long intergenic noncoding RNA (lincRNA) in *Bos taurus*

Tissues used	Number of lncRNA genes/ transcripts detection	Number of analyzed samples	Reference
Mammary gland	184	5	Tong et al. (2017)
Spleen, cortex, adipose tissue, lung, spleen	7235	2	Kern et al. (2018)
Mammary gland	6450	6	Cai et al. (2018)
Milk exosomes	3481	24	Zeng et al. (2019)
Sperm	11,561	6	Wang et al. (2019)

Table 7.4 Genome-wide scale long noncoding RNA (lncRNA) and long intergenic noncoding RNA (lincRNA) in *Gallus gallus*

Tissues used	Number of lncRNA genes/ transcripts detection	Number of analyzed samples	Reference
Liver and adipose tissue	2193	16	Muret et al. (2017)
Vein and skin	15,412	60	Cao et al. (2017)
Ovary	8691	10	Peng et al. (2018)
Spleens	2819	17	You et al. (2019)
Liver	4127	3	Xu et al. (2019)
Uterus, isthmus, ovary	6832	40	Yin et al. (2020)

the expression of lncRNA in different tissues related to growth muscle (Muret et al. 2017). Furthermore, there is a lncRNA expression study that focuses on the development of certain organs, such as the liver (Xu et al. 2019). In the context of female reproduction, fertility was studied (Peng et al. 2018; Yin et al. 2020) (Table 7.4).

7.4.3 lncRNA in Sus scrofa

The Pig lncRNANet database (lnc.rnanet.org) was created by Liang et al. which stores lncRNA findings as well as other published studies including the analysis of expression of lncRNAs in specific organs, such as testis (Ran et al. 2016), liver (Yu et al. 2017; Shen et al. 2018; Li et al. 2016), adipose tissue (Weng et al. 2017), and fat (Table 7.5).

Table 7.5 Genome-wide scale long noncoding RNA (lncRNA) and long intergenic noncoding RNA (lincRNA) in *Sus scrofa*

Tissues used	Number of lncRNA genes/ transcripts detection	Number of analyzed samples	References
Endometrium, fat, embryo, liver, and lungs	7618	5	Li et al. (2016)
Testis	752	6	Ran et al. (2016)
Liver, muscle, and fat	4868	6	Yu et al. (2017)
Adipose tissue	343	6	Weng et al. (2017)
Liver	3368	16	Shen et al. (2018)
Liver	713	3	
Fat	6808	16	

7.5 Metabolomic Biomarkers

Metabolomics is the universal assessment of small endogenous metabolites inside a biological sample utilizing quick and high-throughput methodologies. It is an "omics" field in systems biology and functional genomics. The presence or lack of specific combinations of metabolites is directly connected with the presence or absence of major dietary qualities such as flavor and nutritional value. Metabolite profiling is a method for identifying and quantifying metabolites; however, it has methodological constraints and analytical platform variances. Metabonomics is a subset of metabolomics that studies how organisms respond metabolically to pathogenic stimuli or genetic change. The metabolic investigations could help to better understand the biological processes that cause genetic differences in animals, which is a novel application in animal breeding that combines physiology and genetics. Sundekilde et al. (2011) recently discovered that carnitine, choline, and citrate might be used as possible biomarkers in NMR-based metabolomics to identify the milk of two different dairy breeds (Danish Holstein and Jersey). Boehmer et al. (2010) stated that metabotypes on Holstein milk revealed genome-wide significant relationships with eight metabolites and 21 chromosome-wide significant associations with 14 metabolites. Genome-wide association studies with metabotypes (mGWAS) have linked genomic variability with metabotype levels in relevant biofluids in cattle and pigs. Metabolic hormones and metabolic substrates have a major influence on reproduction in female cattle. Metabolomics is now being applied to diagnostic purposes (Lamers et al. 2005). In canine liver illness, metabolomic studies have been employed to survey metabolic abnormalities (Whitfield et al. 2005). It is successfully used to filter out biomarkers for milk quality (Sun et al. 2015), energy metabolism (Tian et al. 2015), and rumen health in dairy cows (Tian et al. 2016). Metabolomics is a useful tool for analyzing the changes of metabolites in physiological fluids and tissues in response to internal and external stimulations. Yue et al. (2020) analyze the use of diagnostic Biomarkers and Metabolic Changes in Heat-Stressed dairy Cows using a Metabolomics approach. Heat stress (HS) exerts its

impact on physiological parameters, dry matter intake, milk production, the metabolome of milk, and blood plasma in lactating Holstein dairy cows. Heat Stress (HS) lowered dry matter intake (DMI) and affected the metabolites in dairy cows' milk and plasma. HS affected the metabolites in milk as well as the metabolites in plasma. Gluconeogenesis, protein breakdown and synthesis, and milk fat production were the key functions of these metabolites. These compounds found in milk and plasma could be used as indicators or biomarkers of Heat Stress (HS).

Three major approaches used in metabolomics:

1. The targeted analysis states the precise and quantitative measurement of the concentration of a limited number of known metabolites.
2. Metabolic profiling in which untargeted high-throughput measurement of the levels of a large number of metabolites, including unknown metabolites.
3. Metabolic fingerprinting states the quick, total evaluation of biochemical fingerprints for discriminating of distinct groups without the need for metabolite identification.

7.6 Metabolomic Biomarkers for Metabolic Disease Detection in Livestock

Metabolomics is widely employed in a variety of sectors, including pharmacology, toxicology, and diagnostics, and its utilization and technological progress have accelerated. Changes in metabolites are seen in diseased animals as a primary indicator and an important part of clinical practice.

1. *Anterioventral displacement or abomasal displacement (AD)*: AD is a gastrointestinal condition that affects high-producing dairy cows. Hypomotility of the abomasum caused by high concentrate feed and other metabolic disorders, as well as downer cows, are the main causes of abomasal displacement. In the blood, the amino acids valine, 3-hydroxybutyrate, alanine, glutamine, and glutamate are raised, while succinate is decreased.
2. *Ketosis*: It is one of the most common metabolic disorders in dairy cows. During early lactation, high levels of ketone bodies such as hydroxybutyric acid (BHBA), acetoacetate (AcAc), and acetone (Ac), blood glucose levels (Glc), total triglycerides (TG), nonesterified fatty acids (NEFA) such as palmitic acid (PA), heptadecanoic acid (HA), stearic acid (SA), trans-9-octadecenoic acid (T-9-OA), myristic acid (MA), cis-9-hexadecenoic acid (C-9-HA), etc. in the blood, milk, and urine are indicative of ketosis in dairy cattle (Zhang et al. 2021). Reduced levels of leucine (3-hydroxyisovaleric acid (3HIV)), 4-gamma aminobutyric acid (GABA)—L-glutamic acid erythritol, fructose 6-phosphate, and l-serine (L-ser) (Zhang et al. 2012) are evident. Metabolites/lipids including PC ae C30:2, C5-M-DC, PC aa C40:4, and SDMA were selected as biomarkers to characterize ketosis.
3. *Milk fever*: It is a complicated metabolic condition that affects heavy-producing dairy cows around the time of parturition. The metabolites serum amyloid A

protein (SAA), calcitonin gene-related peptide (CGRP), Endopin 2B, Serpin peptidase inhibitors (SPI), Downregulated proteins—fibrinogen beta chain, IGg heavy-chain C-region (IGg-CH), and albumin are used to diagnose milk fever.

4. *Cow Mastitis*: Metabolomics, as the most downstream of expression of genes, can amplify small changes in gene and protein expression at the level of metabolites, providing a fuller understanding of cell function. The use of metabolomics technology in cow mastitis can analyze heterometabolites, identify related biomarkers, and reveal physiological and pathological changes in the cow mammary gland, providing information for mastitis prediction, diagnosis, and treatment (Hu et al. 2021).

7.7 Molecular Biomarkers

Any feature that is objectively measured and evaluated as an indicator of biological processes is referred to as a marker (Mishra and Verma 2010).

The key goal of a breeder is to pick animals with the superior genetic potential to serve as parents for the following generations. Multiple genes or genetic markers associated with genes that affect livestock traits of interest, such as single-gene traits and QTL or genomic regions that affect quantitative traits, have been identified. Molecular markers (direct or indirect ones) have been adopted to screen the individuals with desirable (economic trait loci) or deleterious (disease carriers) alleles through marker-assisted selection (MAS). However, MAS also has several limitations (discussed later), which can be addressed by genomic selection that utilizes genome-wide association study (GWAS).

It is also called a DNA marker. It is a segment of DNA that represents mutations or variations which can be used to identify variation between different genotypes or variants of a gene in a given population or gene pool. The level, structure, and origin of genetic variation can all be studied using molecular markers. Molecular markers allow the detection of variations or polymorphisms.

In gene mapping, molecular markers serve three purposes:

1. It allows direct identification of the gene of interest and makes it a useful tool for screening somatic cell hybrids.
2. It can be found in a variety of DNA probes and also aids in the physical mapping of genes using in situ hybridization.
3. The markers are useful for linkage analysis to generate genetic maps.

7.7.1 Amplified Fragment Length Polymorphism (AFLP)

AFLP technology has been widely used in the identification of genetic polymorphisms, the evaluation and characterization of breed resources, the measurement of breed correlation, the construction of genetic maps, and the identification of genes in

major farm animal species (Marchi et al. 2006). Amplified fragment length polymorphisms (AFLPs) are a useful tool for genotyping, especially when little is known about an organism's genome or genetics. Restriction enzymes cut the DNA and attach adaptors to the fragments' ends. Fragments are then amplified using PCR and their varying lengths can then be visualized on gel or capillary-based platforms. AFLP is very sensitive for detecting genetic polymorphisms but requires relatively large amounts of high-quality DNA and has difficulty with mixture analysis. The AFLP method for DNA fingerprinting was created by combining the techniques of RFLP and PCR. Because of its high multiplex ratio and reproducibility, AFLP has a high power of marker detection. The use of AFLP markers to estimate genetic relationships between individual pigs was also demonstrated by comparing similarity and coancestry coefficient matrices (Ovilo et al. 2000). AFLP markers could be used in MAS for clinical mastitis resistance. The AFLP is used to locate genomic areas containing quantitative trait loci using selected DNA pools.

7.7.2 Single-Nucleotide Polymorphisms (SNP)

A single-nucleotide polymorphism is a difference in DNA sequence caused by a nucleotide substitution at a specific site in the genome. To put it another way, an SNP marker is a single-base replacement in a DNA sequence. SNPs account for more than 90% of all individual differences, making them a useful genetic variation resource for population studies and genome mapping (Frohlich et al. 2004). When compared to the highly informative multiallelic microsatellites, SNP markers are one of the most prominent methods, although they are a step backward (simple biallelic codominant markers). SNPs are a relatively new concept that emerged from the necessity for very high concentrations of genetic markers in the study of complex disorders. SNPs work based on hybridizing identified DNA fragments with high-density DNA probe arrays (also known as SNP-chips), and then naming the SNP allele based on the hybridization results (Yang et al. 2013). SNPs are a type of molecular marker that was developed in the third generation.

SNP markers are a promising useful technology for genetic selection in genomic selection (Seidel 2009). Due to their prevalence in the genome of any organism (coding and noncoding regions) and capacity to uncover hidden polymorphism that is not usually recognized by other markers, these types of markers are becoming increasingly popular in molecular marker development. SNPs with a higher possibility of being in a coding region and so having an effect on phenotypes. In chickens, about a thousand SNPs have been discovered. Identification of particular genes is associated with milk yield and composition, beef composition, lactation persistence, sensitivity to nutritional status, male and female fertility, and other features. Aside from that, SNPs are commonly used in animal disease resistance selection.

7.7.3 *Genome-Wide SNP Detection Through SNP-Chip (Table 7.6)*

7.7.3.1 Bovine SNP-Chips

Illumina has developed SNP-Chips of varying densities, such as the BovineHD Beadchip, BovineLD Beadchip, and BovineSNP50 Genotyping Beadchip, in collaboration with USDA-ARS, University of Missouri, French National Institute for Agricultural Research (INRA), National Association of Livestock and Artificial Insemination Cooperatives, and other frontiers in this area. These Illumina products have a call rate of 99.9% and reproducibility of 99.9% (http://www.illumina.com/products/bovinehd whole-genome genotyping kits.ilmn). The chips also provide holandric and mitochondrial SNPs, which can be used to identify subspecies and paternal and maternal breed lineages (Boichard et al. 2012). Divergent breeds (European Holstein, Jersey, Brown Swiss, Swedish Red and White, Charolais, Limousine, Brahman, N'Dama, Santa Getrudis, etc.) were used to estimate the reference minor allele frequency (MAF) (of North America, Europe, and Oceania). There is a serious concern with the SNP. The lack of flanking information at the end of each chromosome, which contributes to decreased imputation effectiveness in preliminary tests, is a significant issue in terms of SNP density. To address this issue,

Table 7.6 Availability of SNP-chip for genomic selection in different domestic species

Species	Genome assembly	GWAS	SNP-chip	References
Cattle	+	+	• Illumina Golden Gate Bovine3K BeadChip (Bovine3k): 2900 SNPs • Illumina Infinium BovineLD BeadChip (BovineLD): 6909 SNPs • Illumina Infinium BovineSNP50 v.1 BeadChip (BovineSNP50v.1): 54001 SNPs • Illumina Infinium BovineSNP50 v.2 BeadChip (BovineSNP50v.2): 54609 SNPs • Illumina Infinium BovineHD BeadChip (BovineHD): 777962 SNPs • Affymetrix Axiom Genome-Wide BOS 1 Bovine Array (Axiom Bos 1): 648,875 SNPs.	Nicolazzi et al. (2014)
Buffalo	+	+	• Affymetrix Axiom® Buffalo Genotyping Array: 90 K	Brew (2014)
Chicken	+	+	• 600 K Affymetrix® Axiom® HD chicken genotyping array: 580,954 SNPs	Kranis et al. (2013)
Horse	+	+	• Illumina Equine SNP50 Genotyping Infinium Beadchip: 54 K	
Pig	+	+	• Illumina PorcineSNP60 v2 Genotyping BeadChip: 64 K SNPs	Ramos et al. (2009)
Goat	+	+	• Caprine 52 K SNP-Chip: 52,295 SNPs	Tosser-Klopp et al. (2014)

the SNP density in the start and end portions of each chromosome was increased during the construction of the BovineLD BeadChip (Mukhopadhyay and Kumar 2013).

7.7.3.2 BovineLD BeadChip (Illumina)

The low-density BeadChip came into existence after a rigorous work exercised on the Illumina GoldenGate Bovine3K Genotyping Beadchip (http://www.illumina.com/documents/products/datasheets/datasheet_bovine3K.pdf). The major drawback associated with this chip was that the imputation accuracy was compromised due to the dependence of the genotyped individual on the reference population. The quality of the provided genotypes in some samples and the SNP call rate of this 3 K bovine chip were compromised as compared to more advanced chips, like BovineSNP50 chip because GoldenGate chemistry relies on two hybridization events for proper SNP detection as opposed to a single event for Infinium chemistry. In order to cover this gap, the Illumina Infinium BovineLD Genotyping Beadchip (http://www.illumina.com/documents/products/datasheets/datasheet_bovineLD.pdf) was developed, which had high imputation accuracy for higher density SNP genotypes in dairy and beef cattle.

SNPs from the nuclear genome (including sex chromosomes) and mitochondrial genome were included in the Illumina BovineLD BeadChip. It was created to include SNPs from a variety of distant breeds of Bos taurus, B. indicus, and their crosses that are kept in various parts of the world. Except for the telomeric regions, the SNPs in the chip have a high minor allele frequency and uniform spacing across the genome. As a result, in milk and meat cattle breeds, it supports imputation to higher density genotypes. The total number of SNPs included in the low-density chip are 6909. Accuracy of imputation to Illumina BovineSNP50 genotypes using the BovineLD chip was over 97% for most dairy and beef populations (Boichard et al. 2012).

7.7.3.3 BovineSNP50K Chip (Illumina)

The BovineSNP50 chip, with a total of 54,609 SNPs, was produced from known and validated SNPs. The selection priority of SNP was given to following areas (Boichard et al. 2012):

- High MAFs in the breeds in the offset
- Uniform spacing of SNPs (on an average 500 kbp)
- Maintaining the quality of SNP and high reproducibility
- Inclusion of Y-chromosomal and mitochondrial SNPs for determination of sex, parentage, Y haplotypes vis-à-vis subspecies, and maternal lineages

It has the ability to execute up to 24 samples simultaneously. It has a mean gap of 49.4 kb and covers uniformly distributed polymorphic SNPs. The BovineSNP50K chip includes common SNPs that have been validated in economically significant beef and dairy breeds. It has a 0.25 minor allele frequency on average across all loci. The average probe spacing in cattle is 49.4 kb, which provides enough SNP density for reliable genome-association research (Mukhopadhyay and Kumar 2013).

7.7.3.4 BovineHD BeadChip (Illumina)

SNPs validated in economically important beef and dairy cattle are covered by the BovineHD BeadChip, which includes tropically and temporally adapted breeds of Bos taurus, B. indicus, and their hybrids. The average minor allele frequency (MAF) for nonhumped cattle is 0.25 and 0.17 for humped cattle across all loci. With a MAF greater than 0.05, more than 7,49,000 SNPs were validated across all breeds. Over 99% of the markers have been mapped to the UMD3 bovine genome assembly, which includes autosomal, mitochondrial, and sex-linked SNP coverage. With an average gap of 3.43 kb and a median gap of 2.68 kb, uniform genomic coverage provides good SNP density for robust genome-association studies and CNV detection in cattle. The BeadChip can process up to eight samples simultaneously. The BovineHD BeadChip is the most comprehensive genome-wide bovine genotyping array covering near a million of SNPs spanned over the entire nuclear and mitochondrial genome (Mukhopadhyay and Kumar 2013). It can be used to identify the genetic variation across worldwide dairy and beef breeds of cattle for genomic selection, identification of quantitative trait loci (QTL), crossbreed mapping, linkage disequilibrium studies, comparative genetic studies, and breed characterization for evaluating biodiversity.

7.8 High-Throughput Methods

7.8.1 TaqMan Assay or Allele-Specific Hybridization

The TaqMan assay uses Taq DNA polymerase's intrinsic 5′ nuclease activity to generate a fluorescence signal from a short allele-specific oligonucleotide probe. When the probe hybridizes to the PCR template, however, Taq DNA polymerase's 5′ exonuclease activity digests it. The 5′ nuclease assay in which the allelic discrimination is based on Taq DNA polymerase's distinctive 5′–3′ exonuclease activity (Holland et al. 1991; Livak et al. 1995). As a result, the two fluorophores are in solution, and the quenching effect is eliminated. The allelic discrimination is based on the 5′ to 3′ characteristic.

7.8.2 Microarray or Gene Chips or DNA Chips

During the last two decades, the most widely used technology for monitoring the cellular abundances of transcript species has been DNA chips. Hybridization and fluorescence detection are used to distinguish alleles. SNPs can also be detected using DNA chips and microarrays of immobilized oligonucleotides of known sequences that differ at specified places of the SNP site. To improve genotyping accuracy, several probes varying at a single site are utilized to analyze each SNP. It is a rapid, reliable, and cost-effective technique. Due to the use of linkage disequilibrium (LD) phenomenon, the microarrays are designed to describe genetic variation within a genome of interest in the best feasible way.

7.8.3 Pyrosequencing

It is a new rapid resequencing technique in which pyrophosphate (PPi) release is used to track the late-mediated, oligonucleotide-primed incorporation of nucleotides by a polymerase. The light is produced when luciferin is oxidized by luciferase. The energy for this reaction comes from pyrophosphate to adenosine triphosphate conversion (ATP). This technology has significant advantages in terms of speed and accuracy, and it does not require the use of electricity. However, depending on the local sequence, designing suitable primers is challenging, and an apparatus specifically developed for this study is needed.

7.8.4 Oligonucleotide Ligation Assay (OLA)

The activity of the DNA ligase enzyme is used to ligate two specific primers that are close to each other in this experiment. On one side of the SNP, two oligonucleotides specific for each allele and labeled differently are designed, and on the other side, one common oligonucleotide is designed (Tobe et al. 1996). When hybridized to the complementary target DNA sequence in question, two primers are created that are directly next to each other for the oligonucleotide ligation assay. To be covalently bonded by ligation, the two adjacent primers must be immediately next to each other with no gaps or mismatches.

7.8.5 Flap Probe Cleavage Approach

A new chemical method for detecting single bases was recently described. This strategy is based on the discovery that ancient flap endonucleases also known as

cleavages recognize and cleave a structure generated when two overlapping oligonucleotides hybridize to a complementary DNA target. When the fluorescence resonance energy transfer (FRET) probe is cleaved, the final detection can be relied on the variations in fluorescence intensity (Lyamichev et al. 1999).

The ability to avoid PCR amplification, the inexpensive cost of unlabeled allele-specific probes, and the relative simplicity of the reaction technique are the key advantages of this approach. The disadvantage is that the intruder assay requires a considerable volume of genomic DNA when PCR is not utilized.

7.8.6 Microsatellite (SSR)

Microsatellites markers are easy to analyze, highly polymorphic DNA markers. Microsatellites are DNA loci that include 1–6-nucleotide repeats that are aligned without being interrupted by any other base or motif and are scattered throughout the genome. Short tandem repeats (STRs), simple sequence repeats (SSRs), and simple sequence tandem repeats (SSTR) are all synonyms of microsatellites loci. They are DNA segments with a variable number of copies (typically 5–50) of two to five (2–5) base sequence motifs known as a repeat unit. They are polymorphic and abundant, and they are frequently located in noncoding sections of genes. Molecular characterization aids in the genetic management of small populations to prevent inbreeding. Molecular markers provide the foundation for enhancing conservation efforts (Hanotte and Jianlin 2005). Because of their high mutation rate, microsatellites are the most informative molecular marker, with the advantage of facile and low-cost PCR detection. Microsatellites also have the advantage of being codominant, which means they detect both homozygote and heterozygote genotypes, unlike RAPD and AFLP markers, which only detect the presence or absence of a locus.

7.8.7 Random Amplification of Polymorphic DNA (RAPD)

The approach is based on the amplification of genomic DNA with either a single or several short oligonucleotide primers of an arbitrary or random sequence, commonly known as arbitrarily primed polymerase chain reaction (AP-PCR). Intraspecific variation detected using RAPD primers differs per species. The RAPD-PCR method was proven to be successful in finding polymorphisms in bovines. Because molecular markers are unaffected by selection processes, they can be used to distinguish closely related genotypes. The use of amplified DNA sequences as molecular markers was first introduced by the polymerase chain reaction (PCR) technology. The RAPD-PCR method is used to discover relationships between different breeds and individuals, formulate breeding strategies, and control animal pedigree data (Talle et al. 2005). Kemp and Teale (1994) employed the RAPD approach to successfully define bovine herds. Other studies that used RAPD markers include

those conducted on Japanese black cattle (Wagyu), Zebu cattle, German native cattle, and Korean native cattle (Yeo et al. 2000). Hwang et al. (2001) used RAPD markers in their linkage analysis of poultry. For Zimbabwe, RAPD is utilized to build a conservation and breeding program. As a result, identifying RAPD primers that can be used to determine the genotypes in cattle is critical. The key advantages of the RAPD technique over other DNA marker assays are that it is easier, less labor-intensive, faster, and less expensive.

7.8.8 Restriction Fragment Length Polymorphism (RFLP)

It is a common tool for checking small but unique alterations in a double-stranded DNA sequence. It is based on the specificity of restriction endonucleases, which recognize and cleave DNA at a set of nucleotides known as restriction sites. RFLPs are commonly utilized in the study of diversity and phylogenetic relationships among closely related species, as well as in the creation of genetic maps.

Because of the abundant availability of diverse restriction enzymes and their random distribution throughout the genome, RFLPs have been routinely used in gene mapping investigations.

It is used to genetically distinguish between organisms by analyzing distinctive patterns in DNA fragments. Polymerase Chain Reaction and Restriction Fragment Length Polymorphism (PCR-RFLP) analysis was used to differentiate variations in two indigenous Bos indicus breeds, the Sahiwal and Tharparkar cattle (Rachagani et al. 2006) (Tables 7.7 and 7.8).

Merits of Biomarkers

1. Objective evaluation.
2. Risk or disease homogeneity.
3. Trustworthy; validity may be proven.
4. Disease processes are frequently researched.
5. They are less skewed than questionnaires.
6. Measurement accuracy.

Table 7.7 Different characteristics of molecular markers

Molecular markers	Levels of polymorphisms	Efforts to genotype	Cost and labors	Reliability	Accuracy
RFLP	Low	High	Low	Very high	Very high
PCR-RFLP	Low	Medium	Medium	High	Very high
RAPID	Medium	Very low	Medium	Low	Very low
AFLP	Medium	Very low	Medium	High	High
SSR	High	Low	High	High	High
SNP	High	Variable	High	High	Very high

Table 7.8 Advantages and disadvantages of molecular markers

Type of markers	Advantages	Disadvantages
Random amplification of polymorphic DNA (RAPD)	• There is a huge amount of it in the genome • A less amount of DNA was required • It takes less time • For detection, there is no radioactive labeling • Genome coverage is extensive • There is no requirement for sequence information	• There is no information on the probe or markers • It is not possible to use it across species
Restriction fragment length polymorphism (RFLP)	• Coverage of the entire genome • The genome contains a lot of it • There is no requirement for sequence information • High reproducibility can be used for a variety of species	• It is difficult to automate this process • More time-consuming • It takes a long time • A huge amount of high-quality DNA is required • For detection, radioactive labeling is required
Amplified fragment length polymorphism (AFLP)	• Highly reproducible • Provides good genome coverage and is very repeatable • Extremely polymorphic • Used to store sequence information for any organism	• Complicated methodology • High quality and quantity of DNA required
Simple sequence repeat (SSR) or microsatellite	• Simple automation • For detection, there is no radioactive labeling • Coverage of the genome across a medium distance	• It is not possible to use it across species • Sequence information is required
SNP (Single-nucleotide polymorphism)	• Cost-effective • Codominant marker • No need for prior sequence knowledge	• High development cost

Limitations of Biomarkers

1. It is crucial to be on time.
2. Mistakes in the laboratory.
3. Safekeeping (longevity of samples).
4. Establishing a normal range is tough.
5. Responsibilities in terms of ethics.
6. Exorbitant (costs for analyses).

7.8.9 Genome-Wide Association Study (GWAS) as Powerful Markers in Livestock History

In order to discover positional candidate genes for the variables under investigation, the GWAS examines all identified markers (about 1 M apart) that are spread over the whole genomes received from roughly 1000 or more individuals in each of the treatment and control groups.

As a result, GWAS aids in the identification of candidate genes, which are subsequently utilized to look for causative mutations for the traits under investigation. The GWAS can identify single-nucleotide polymorphism (SNP) and other DNA variants that are linked to a trait, but it cannot pinpoint the causal genes on its own.

The Human Genome Project was completed in 2003 and the International HapMap Project was completed in 2005 (International HapMap Project, 2005).

The GWAS has been successfully extended toward animal science, for detecting the differentially expressed genes as well as identification of the central key gene (s) underlying the trait(s) of interest, followed by construction of network. It encompasses disease tolerance/susceptibility, production vis-à-vis reproduction traits and growth traits, as well. The intermediate phenotypes (viz. enzyme concentration, fatty acid content in meat, specific casein content in milk, etc.) are measured (Danesh and Pepys 2009) and subjected to analysis for determining the gene clusters contributing to the traits or detecting the differentially expressed genes in two groups. Presently, several approaches are followed to find out the differentially expressed genes/transcripts between the two groups, namely, cluster analysis (Eisen et al. 1998), weighted gene co-expression network analysis (WGCNA) (Zhang and Horvath 2005), partial correlation information theory (PCIT) (Reverter and Chan 2008), etc.

7.8.10 Significance of GWAS in Molecular Animal Breeding

- To associate between the variations in genotypes and phenotypes to identify the causal genetic variants
- To identify QTL underlying many common, complex disease governing the biological pathways
- To associate a trait with a region in the genome, in order to map the clinically and/or economically important QTLs

7.9 Genomic Selection (GS)

"Exploit dense marker data to image the genome, link genotype and phenotype information for estimating marker effects, and lastly use genomic breeding values (GEBV) as aggregates of all marker effects for selection," according to genomic selection (GS).

Unlike traditional pedigree-based evaluations, the GS predicts molecular breeding values using marker genotypes from the complete genome. In a nutshell, genome-wide marker effects are predicted using marker data and phenotypic data using a suitable statistical model (i.e., Bayesian technique).

The genomic breeding value (GEBV) of the animals whose SNPs were employed is estimated using the impacts of these SNP genotypes. The accuracy and reliability of the prediction model are next evaluated. If the prediction equation is determined to be suitable, it is applied to closely related populations using only marker data, allowing future breeding animals to be picked at an early age.

Before discussing the advantages and disadvantages of GS, a brief history of traditional breeding and MAS will be reviewed, with the goal of shedding light on the paradigm change from traditional breeding to molecular breeding.

7.10 Genomic Selection in Dairy Cattle

The majority of the early findings on genetic selection used simulated data. These studies indicated that GS can be used to maximize genetic gain and solve the constraints of MAS in domestic animals, particularly large animals with longer generation intervals.

Later, GS studies were carried out in a range of domestic animals, either to choose breeding animals or to explore for genome-wide connections with economically important characteristics. Some countries, including as the United States, New Zealand, Australia, and the Netherlands, have already assessed the dependability of predicted GEBV in dairy cattle. Reference populations of 650 to 4500 progeny-tested Holstein-Friesian bulls were genotyped for roughly 50,000 genome-wide markers in these studies. In the reference population, GEBV reliability for young bulls without progeny test results ranged from 20% to 67%. In Canada, a GS project has been launched to genotype 1000 Holstein-Friesian dairy sires in order to produce a reference population for "training" the SNP-chip. Technology platforms are now available to examine the variation among animals for 54,001 SNPs. In the Canadian dairy population, a cost-benefit analysis suggests that double genetic gain at 92% of the cost. At North Carolina State University, Gray et al (2011) investigated the feasibility of using genomic selection by establishing the efficiency of genomic prediction of milk flow features in the Italian Brown Swiss population (USA). In most situations, the genetic worth for milk flow features determined using genomic markers was more reliable than traditional pedigree-based analyses.

Table 7.9 Biomarker discovery in current technologies: advantages and disadvantages

Biomarkers	Advantages	Disadvantages
Proteins and peptides	Protein identification with high specificity and accuracy	Significant optimization is required, it is time demanding, it has a restricted dynamic range of detection, and it is impacted by abundant proteins
	High throughput, absolute measurement, small sample sizes, and no depletion of abundant proteins are all advantages of this method	Potential cross-reactivity of antibodies or aptamers may contribute to false positives if detection is limited to certain protein targets
Circulating immune cells	High throughput, with the ability to screen numerous markers at the same time	Due to spectral overlap, there are a limited amount of markers
ctDNA	High accuracy and reproducibility at a low cost, sensitivity, precision, and repeatability	There is no uniformity, and each test is limited to 1–2 mutations High cost, high complexity, and low sensitivity
Circulating tumor cells	Processing time is quick, and the price is reasonable A standard curve is required, and primer design affects specificity	It necessitates a big sample volume
miRNA and lncRNA	It is widely utilized, simple to use, and inexpensive	A standard curve is required, and specificity is determined by primer design

Female reproductive parameters of Holstein Friesian (HF) cows were assessed across the country in Ireland, the United Kingdom, the Netherlands, and Sweden. For bivariate genome-wide associations of traditional fertility parameters (viz. days to first service, days to first heat, pregnancy rate to first service, number of services, and calving interval) and fertility phenotype derived from milk progesterone profiles, Bayesian stochastic search variable selection using Gibbs sampling was used. It was concluded that exchanging data rather than relying on physiological markers of the trait under inquiry could boost the GWAS's power. At Trinity College Dublin, Ireland, researchers looked at the composite phenotype of genetic merit for tuberculosis susceptibility among the daughters of HF top sires (Table 7.9).

7.10.1 Next-Generation Sequencing (NGS) in Livestock

Sanger sequencing by dideoxy chain-termination method had been the central approach to determine DNA sequence in Animal Biotechnology for more than 30 years and was regarded as the gold standard for addressing genomic variants before next-generation sequencing (NGS) came. NGS is a word used to describe post-Sanger sequencing technology and the chemistry and technique employed in the sequencing process are referred to known as "generation." It originally appeared in 1977 and then dominated for the next three decades (Mardis 2008). Deep, high-

throughput, in-parallel DNA sequencing technologies are referred to as next-generation sequencing (NGS) or "High-Throughput Sequencing Technologies." NGS technology can sequence millions to billions of reads in parallel in a single run, and the time it takes to generate GigaBase-sized reads is only a few days or hours, making it superior to first-generation sequencing methods like Sanger sequencing or chain-termination method. Sanger et al. (1977) invented and published a method for "DNA sequencing using chain-terminating inhibitors" that used chain-terminating dideoxynucleotide analogs to trigger base-specific termination of primed DNA synthesis. These methods can be utilized for whole-transcriptome shotgun sequencing (WTSS)—also known as RNA sequencing (RNA-seq) (Wang et al. 2009), whole-exome sequencing (WES) (Rabbini et al. 2014), targeted (TS), or candidate gene sequencing (CGS) (Leo et al. 2015), and methylation sequencing (MLS) (Pelizzola and Ecker 2011) in addition to WGS. DNA sequencing technology should ideally be quick, accurate, simple to use, and inexpensive. NGS technology has an inclusive impact because it allows small and large research groups to provide answers and solutions to a wide range of problems and questions in biology and genetics, including those in medicine, molecular cloning, breeding, discovering harmful genes, comparative or evolutionary studies, cattle ranching, forensic science, parasitology, microbiology, marine and plant biology. Such panels helped in livestock with precise mapping of genetic damage (Berry et al. 2011), characterization of quantitative trait loci, genetic study of population structure (Edea et al. 2013), and species evolutionary history (Alföldi and Lindblad-Toh 2013). A wide spectrum of "omics," including genomics, transcriptomics, and epigenomics, has been characterized and quantified. Omics is essential to understand the mechanisms and functions of many molecules.

The most essential aspect, however, was the incorporation of molecular technologies into animal breeding to increase breeding value. Traditional animal breeding is considerably enhanced by the genomic selection, which complements phenotypic selection procedures (Dekkers 2012) (Table 7.10).

7.11 Different Platforms of NGS

7.11.1 First-Generation Sequencing

The first natural polynucleotide sequence was reported 12 years after Watson and Crick published their double-helix DNA structure in 1953 (Holley et al. 1965). The phiX174 genome, which has a size of 5374 bp, was the first genome sequenced using Sanger sequencing (Sanger and Coulson 1975).

Table 7.10 Various NGS online databases and sites

Tools	Website
The NGS WikiBook	http://en.wikibooks.org/wiki/Next_Generation_Sequencing_(NGS)
The Sequencing Marketplace	http://allseq.com
Genomeweb	https://www.genomeweb.com
Bioinformatic software	http://seqanswers.com/wiki/Software/list
Bioinformatics Web	http://www.bioinformaticsweb.net
Biological databases	https://en.wikipedia.org/wiki/List_of_biological_databases
Applied Bioinformatics	http://www.appliedbioinformatics.com.au
dbSNP	http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi
dbGAP	http://www.ncbi.nlm.nih.gov/gap
Complete Genomics data	http://www.completegenomics.com/public-data/
SRA	http://www.ncbi.nlm.nih.gov/sra
OMIM	http://www.ncbi.nlm.nih.gov/omim
COSMIC	http://cancer.sanger.ac.uk/cosmic
ENCODE	https://www.encodeproject.org
GTEx	http://www.gtexportal.org
FANTOM	http://fantom.gsc.riken.jp
Roadmap epigenomics	http://www.roadmapepigenomics.org
Blueprint epigenomics	http://www.blueprint-epigenome.eu
Regulome DB	http://regulomedb.org
ExPASy proteomics	http://www.expasy.org/proteomics/protein-protein_interaction
PRIDE proteomics	http://www.ebi.ac.uk/pride/archive/
FAME metabolomics	http://f-a-m-e.fame-vu.cloudlet.sara.nl
MetabolomeExpress	https://www.metabolome-express.org
MetaboAnalyst	http://www.metaboanalyst.ca
AromaDeg	http://aromadeg.siona.helmholtz-hzi.de
EGA phenome	https://www.ebi.ac.uk/ega/home
GOLD	https://gold.jgi-psf.org
MG-RAST	https://metagenomics.anl.gov
ViralZone	http://viralzone.expasy.org
UCNEbase UC elements	http://ccg.vital-it.ch/UCNEbase/
DEG database	http://www.essentialgene.org
PhylomeDB	http://phylomedb.org/
Compara GeneTrees	http://asia.ensembl.org
TreeFam	http://treefam.genomics.org.cn
PANTHER	http://pantherdb.org
FATCAT	http://phylogenomics.berkeley.edu
HOGENOM database	http://doua.prabi.fr
Gene Ontology Consortium	http://geneontology.org
OBBO	http://www.obofoundry.org, http://obofoundry.github.io
Reactome	http://www.reactome.org/
DAVID 6.7	https://david.ncifcrf.gov/

(continued)

Table 7.10 (continued)

Tools	Website
Kbase	http://kbase.us/glossary/systems-biology/
Earth Microbiome Project	http://www.earthmicrobiome.org
Terragenome Project	http://www.terragenome.org
Oceans Project	http://ocean-microbiome.embl.de/companion.html
MetaHit project	http://www.metahit.eu
Vertebrate Genome 10 K	http://genome10k.org
Human Microbiome	http://hmpdacc.org
Personal Genome Project	http://www.personalgenomes.org
1000 Genomes Project	http://www.1000genomes.org
HapMap	http://hapmap.ncbi.nlm.nih.gov/
UCSC browser	https://genome.ucsc.edu
Ensembl browser	http://www.ensembl.org
Jbrowse browser	http://jbrowse.org
Web Apollo browser	http://genomearchitect.org
NCBI mapview	http://www.ncbi.nlm.nih.gov/projects/mapview/
NCBI resources	http://www.ncbi.nlm.nih.gov/
KEGG	http://www.genome.jp/kegg/

7.11.1.1 Sanger Sequencing or Chain-Termination Sequencing

It requires using one strand of double-stranded DNA as a sequencing template. Dideoxynucleotides are chemically modified nucleotides that are used in this sequencing (dNTPs). These dNTPs are labeled with ddG, ddA, ddT, and ddC for each DNA base. Dideoxynucleotides (dNTPs) are utilized for nucleotide elongation; once incorporated into the DNA strand, they stop further elongation and the elongation is complete. Then we get DNA fragments with various diameters that are terminated by a dNTP. The pieces are separated by size on a gel slab, with the resulting bands corresponding to DNA fragments visible using imaging techniques (X-ray or UV light) (Masoudi-Nejad et al. 2013; El-Metwally et al. 2014).

Because of its greater simplicity and reliability, as well as the use of fewer toxic chemicals and lower amounts of radioactivity, the new Sanger DNA chain-termination sequencing method (Sanger et al. 1977) emerged victorious over the Maxam and Gilbert chemical degradation method (Maxam and Gilbert 1977).

7.11.1.2 Maxam-Gilbert Chemical-Degradation Sequencing

Allan Maxam and Walter Gilbert reported chemical sequencing, commonly known as chemical degradation, in 1977, which required chemical changes to the DNA as well as additional breakage and electrophoresis (Maxam and Gilbert 1977). Because it permitted direct sequencing of pure DNA without requiring further in vivo cloning

and ssDNA preparation processes, the Maxam-Gilbert sequencing technology was a huge development and became the methodology of choice.

The process involves utilizing polynucleotide kinase to radioactively label the 5′-P ends of double-stranded DNA (dsDNA) with 32P-dATP. The DNA is next denatured at 90 °C with DiMethyl SulfOxide (DMSO), and the ssDNA molecules are separated by electrophoresis.

Adenosine (A), Cytidine (C), Guanosine (G), and Thymidine (T) are modified using nitrogenous base-specific processes' residues, allowing the chemical cleavage of the ssDNA at the 5′-P side of such positions. There is some slight G and C cleavage produced by the A and T reactions, which should be considered (showing as a weaker signal later on). Following that, polyacrylamide gel electrophoresis and autoradiography are used to separate ssDNA fragments by size and identify the radiolabeled DNA band pattern on an x-ray film that encodes the DNA sequence, allowing the sequence to be determined. From 2003 to 2009, the genomes of companion animals such as the cat and the dog as well as farm animals like the cattle, pig, chicken (Hillier et al. 2004), sheep, and horse were sequenced.

7.11.2 Second-Generation Sequencing

The mechanism of second-generation or next-generation sequencing techniques is based on clonal multiplication of the DNA molecule, in which billions of distinct DNA fragments are sequenced in parallel at the same time, generating massive amounts of data. The production and amplification of libraries created from DNA or RNA samples, clonal formation, sequencing, and analysis are all part of the overall workflow for second-generation sequencing (Head et al. 2014).

7.11.2.1 Roche/454 Sequencing

Roche purchased 454 Life Sciences, a pyrosequencing technology-based sequencer, in 2007 and renamed it Roche 454. Each fragment of DNA is connected to a bead whose surface contains primers with oligonucleotides corresponding to the DNA fragments, resulting in each bead being associated with a single fragment. Then, using PCR emulsion, each bead is extracted and amplified, yielding approximately one million copies of each DNA fragment on the surface of the bead. The beads are then placed to a picotiter plate (PTP) with several wells and the pyrosequencing approach is used, which involves initiating a sequence of downstream reactions that produce light at each nucleotide incorporation. The sequence of the DNA fragment is inferred by monitoring light emission after each nucleotide incorporation. Hundreds of thousands of reactions can run in parallel on a picotiter plate, significantly increasing sequencing throughput (Vezzi 2012). Roche/454's latest instrument, the GS FLX+, generates reads up to 1000 bp in length and can produce one million reads per run (454.com GS FLX + Systems http://454.com/products/gs-flx-system/index.

asp). For high-throughput real-time sequencing, this is a fast (700 MB data in a day), accurate (99.9% after the filter), and reliable technique. The Roche 454 is a more advanced platform that can provide read lengths of over 700 bp.

7.11.2.2 Sequencing by Oligonucleotide Ligation and Detection (SOLiD)

SOLiD is a next-generation sequencing device that was first produced in 2008 by Applied Biosystems Instruments (ABI)001 and is now marketed by Life Technologies (http://www.lifetechnologies.com). DNA sequencing approach uses the DNA ligase enzyme's mismatch sensitivity to determine the DNA sequence (Ho et al. 2011). There are five major steps in the sequencing reaction that include DNA library preparation, emulsion PCR clonal amplification in microreactors, bead attachment, sequencing, primer resetting.

Multiple rounds of sequencing are involved in the process. It begins with the addition of adapters to DNA fragments, which are then fixed on beads and cloned using PCR emulsion. The 8-men with a fluorescent label at the end is then successively ligated to DNA fragments on a glass plate, and the color emitted by the label is recorded. The output format is color space, which is the nucleotide's encoded form, with four fluorescent colors representing 16 different combinations of two bases. The sequencer repeats this ligation cycle, removing the complementary strand after each cycle and starting a new sequencing cycle at position n-1 of the template. The cycle continues until each base has been sequenced twice. Because each nucleotide is sequenced twice. This technology has a maximum accuracy of 99.99% (Voelkerding et al. 2009). As a result, there are fewer risks of miscalling from two consecutive colors.

7.11.2.3 Illumina/Solexa Sequencing

A novel method of sequencing has been developed by the Solexa firm. The Illumina firm (http://www.illumina.com) purchased Solexa, which began commercializing the Illumina/Solexa Genome Analyzer sequencer. Illumina systems are the market leaders in high-throughput sequencing. Illumina is currently working on several platforms (MiSeq, HiSeq series, and NextSeq series). The fragmentation of template DNA and end fixing of fragments (blunting and 5′ end phosphorylation) are used in the sequencing approach to enhance ligation with oligonucleotide adapters with a "T" overhang at 3′, adenylation of 3′ ends with the addition of a single "A" nucleotide.

Anchors are hybridized because ligated adapters are complementary to the flow cell. In contrast to emulsion PCR, the DNA template connected to the flow cell anchors relies on "bridge amplification" for cluster creation (Adessi et al. 2000) in the parallel sequencing of millions of clusters onto the flow cell. The distal end of a DNA fragment forms an arc and hybridizes with an adjacent anchor oligonucleotide to form a complementary portion. Each template generates thousands of copies of

the same template DNA due to clonal amplification, and millions of unique clusters are generated on a single flow cell after DNA polymerase is added, and four different fluorescent-labeled reversible terminators aid in sequencing the millions of clusters in parallel onto the flow cell. The fluorescent label is cleaved by enzymatic cleavage, allowing the inclusion of the next nucleotide (www.illumina.com).

7.11.2.4 Ion Torrent Sequencing

The developers of 454 sequencings invented Ion Torrent technology (http://www.iontorrent.com). For example, employs a chip with a series of microwells, each of which holds a bead containing multiple identical fragments. When each nucleotide binds to a fragment in the pearl, a hydrogen ion is produced, causing the pH of the solution to alter.

A sensor attached to the bottom of the microwell detects this change and converts it into a voltage signal proportional to the number of nucleotides integrated. The Ion Torrent sequencers can generate reads with lengths of 200, 400, and 600 bp, with a throughput of up to 10 Gb for the ion proton sequencer. The main benefits of this sequencing method are the greater read lengths compared to other SGS sequencers, flexible workflow, faster turnaround time, and lower pricing.

7.11.3 Third-Generation Sequencing (TGS)

TGS is characterized by two main approaches: the single-molecule real-time sequencing approach (SMRT) developed by Quake laboratory and the synthetic approach, which relies on existing short reads technologies such as those used by Illumina (Moleculo) (Harris et al. 2008) and 10 × Genomics (https://www.10xgenomics.com) to construct long reads. Because of single-molecule and real-time sequencing technologies, third-generation sequencing (TGS) has gained insight into more genetic illnesses in recent years. Pacific Biosciences and Oxford Nanopore sequencing (particularly the MinION sequencer) have used the SMRT technology technique, which is the most extensively used TGS technology approach.

7.11.4 Pacific Biosciences Single-Molecule Real-Time (SMRT) Sequencing

Phospholinked nucleotides and zero-mode waveguides are the foundations of this sequencing method (ZMW). It is the real-time sequencing of single molecules (SMRT). ZMW is unique in that it only allows light to illuminate the bottom of a well where the immobilized template and DNA polymerase are present. It detects the

signals as they are released during the incorporations in real-time. It employs a device made up of multiple SMRT cells, each of which comprises microfabricated nanostructures known as zero-mode waveguides (ZMWs), which are wells with diameters of tens of nanometers microfabricated in a metal film and then deposited onto a glass substrate. These ZMWs take advantage of the qualities of light traveling through apertures that are smaller than its wavelength, preventing light from propagating. The light intensity falls throughout the wells due to their narrow diameter, and the bottom of the wells is illuminated. A DNA polymerase is attached to the bottom of each ZMW, as well as the target DNA fragment for sequencing. The DNA fragments are separated during the sequencing procedure and integrated by the DNA polymerase with fluorescently tagged nucleotides during the sequencing step (with different colors). When a nucleotide is added, it emits a bright signal that is detected by sensors. It is feasible to determine the DNA sequence by detecting the tagged nucleotides. Pacific Biosciences sequencing platforms have a high error rate of around 13% with insertions and deletions being the most common errors.

7.11.5 Oxford Nanophore

Nanopore technology refers to nanoscale holes implanted in a thin membrane structure that are used to detect potential changes when charged biological molecules smaller than the nanopore pass through the hole.. Oxford Nanopore Technologies recently announced the first commercial nanopore sequencing devices, which will be available by the end of 2012, while other companies (Life, Roche, IBM) are also pursuing nanopore sequencing approaches. As a result, nanopore technology has the ability to detect and analyze single-molecule amino acids, DNA, RNA, and other biological molecules. A hairpin connects the initial strand of a DNA molecule to its corresponding strand in this sequencing technique. A protein nanopore (a nanopore is a tiny hole formed of proteins or synthetic materials) is used to transmit the DNA fragment through it. When a motor protein linked to the pore translates a DNA fragment via the pore, it causes an ionic current fluctuation due to changes in the moving nucleotides filling the pore This ionic current variation is progressively recorded on a graphic model and then evaluated to determine the sequence. The direct strand generates the "template read," and the hairpin structure is read, followed by the inverse strand, which generates the "complement read." These reads are referred to as "1D." When the "temple" and "complement" reads are merged, we get a "two direction read" or "2D" consensus sequence (Lu et al. 2016; Jain et al. 2016) (Tables 7.11 and 7.12).

Table 7.11 Comparison of different NGS platforms

Method	Sanger sequencing	SMRT	Ion conductor	Pyrosequencing	Illumina	SOLiD sequencing
Read length	400-900 bp	2900 bp	200 bp	200 bp	50 + 250 bp	50 + 50
Accuracy	99.9	87	98	99.9	98	99.9
Reads per run	NA	35–75 thousand	2 h	1 million	Up to 3 billion	1.2–1.4 billion
Times per run	20 min–2 h	30 min–2 h	2 h	24 h	1–10 days	1–2 weeks
Advantages	Long individual reads	Fast, long individual reads	Less expensive, fast	Long read size, fast	High sequence reads, fast	Low cost per base
Disadvantages	Empirical and more expensive	Low yield	Homopolymer errors	Runs are expensive	Expensive equipments	Quite slower than others

Table 7.12 Details of the first livestock whole-genome sequence assemblies deposited in NCBI

Species	Common name	SRA[a] experiments
Bos taurus	Cattle	12,589
Equus caballus	Horse	3495
Gallus gallus	Chicken	3637
Canis familiaris	Dog	3751
Anas platyrhynchos	Duck	570
Sus scrofa	Pig	6270

[a]Sequence read archive—that stores the raw sequencing data and alignment information from high-throughput sequencing platforms at NCBI

7.12 Status of Domestic Animal RefSeq (Reference Sequencing)

Several domesticated livestock species' genomes (chicken, pig, cow, sheep, and horse) have been partially or entirely sequenced in recent years. Although sex chromosomes were inadequately annotated in the initial assembly, the first draught of the chicken genome was constructed using an assembly with 6.6-fold whole-genome shotgun coverage (Hillier et al. 2004). The Illumina SNP BeadChip for commercial chicken (broilers and layers) was designed with 352,303 SNPs.

The Bovine Genome Sequencing and Analysis Consortium published the Taurine cow genome sequence in April 2009, after it had been sequenced and assembled with roughly sevenfold coverage (Elsik et al. 2009) (Tables 7.13, 7.14, and 7.15).

7.13 Applications of NGS as Biomarkers in Livestock

7.13.1 In Pork Production

The livestock industry may benefit directly from this cycle of biomarker development: as scientific knowledge develops, more and better biomarkers will be identified that can be used by the industry, provided that the developed biomarkers have economic relevance for the industry. The livestock industry may use biomarkers to gain better insight into relevant traits (trait development) to be able to improve the traits in the desired direction. The livestock industry may use biomarkers as determination (diagnostic) tools, monitoring tools, decision making/management tools.

7.13.2 Animal Conservation

Disease resistance and other desirable traits in indigenous breeds could be the result of genetic heterogeneity in the germplasm. The recording of genetic information and

Table 7.13 Status of next-generation sequencing (NGS) in livestock (www.ncbi.nlm.nih.gov)

Animal	Common name	Breed	Sex	Genome size	Conting count	Total sequence length	Platform used	Scaffold count
Bos taurus	Cattle	Hereford	Female	2.86 billion base pair	42,267	2,87,49,56,119	PacBio RSII	5998
Bos indicus		Nelore	Male		253,770	2,673,965,444	SOLiD	32
Equus caballus	Horse	Thoroughbred	Female	2.4–2.7 GB	9687	2,47,49,29,062	NA	9687
Gallus gallus	Chicken	Red jungle fowl	Female	1.2 GB	1602	1,09,87,54,166	PacBio RSII	528
Sus scrofa	Pig	Mixed	Female	2.7 Gb	1118	2,26,25,79,801	PacBio	706
Ovis aries	Sheep	Texel	Female		48,482	2,615,499,683	PacBioRsII	5465
Apis mellifera	Honey bee	DH4	Strain		5645	2,50,287,000	SOLiD, life 454	5645

Table 7.14 Useful tools and websites for NGS assembly and postassembly

Tools	Website
Bowtie aligner	http://bowtie-bio.sourceforge.net/index.shtml
Velvet assembler	https://www.ebi.ac.uk/~zerbino/velvet/
Anytag aligner	http://sourceforge.net/projects/anytag/files/anytag2.0/
Celera assembler	http://wgs-assembler.sourceforge.net/wiki/index.php?
Genomic tools	http://molbiol-tools.ca/Genomics.htm
SPAdes assembler	http://bioinf.spbau.ru/spades
BaseSpace Illumina	https://basespace.illumina.com/home/sequence
MUMmer aligner	http://mummer.sourceforge.net
Galaxy tools	https://usegalaxy.org

Table 7.15 Prokaryotic and eukaryotic annotation online servers and their websites

Server name	Website
Prokka	http://www.vicbioinformatics.com/software.prokka.shtml
RAST	http://www.nmpdr.org/FIG/wiki/view.cgi/FIG/
CRISPRfinder	http://crispr.u-psud.fr/Server/CRISPRfinder.php
Mreps	http://bioinfo.lifl.fr/mreps/mreps.php
MicroScope	https://www.genoscope.cns.fr/agc/microscope/home/index.php
BaSys	https://www.basys.ca
PGAP	http://www.ncbi.nlm.nih.gov/genome/annotation_prok/
NCBI pipeline	http://www.ncbi.nlm.nih.gov/genome/annotation_euk/process/
RepeatMasker	http://www.repeatmasker.org/
Censor	http://www.girinst.org/censor/
WindowMasker	http://nebc.nerc.ac.uk/bioinformatics/docs/windowmasker.html
CEGMA tool	http://korflab.ucdavis.edu/datasets/cegma/
BUSCO	http://busco.ezlab.org
PASA	http://pasapipeline.github.io
MAKER	http://www.yandell-lab.org/software/maker.html
Babelomics	http://www.babelomics.org

the future conservation of local breeds can be aided by whole-genome sequencing of domestic and wild animal species. Unlike PCR-based approaches, NGS is a sequence-independent detection tool that can be used for forensic identification of endangered animals (Marin et al. 2009).

7.13.3 Breeding of Animals

In addition, next-generation sequencing (NGS) has been employed in chicken production in the United States. Genes affect the quality of egg shells. From the economic point of view, broken egg shells result in significant losses and may pose a

health risk to humans. The introduction of next-generation sequencing (NGS) was a watershed moment in molecular biology and animal genetics and breeding.

7.13.4 Microbiological Approach

The next-generation sequencing (NGS) approach allows microorganisms to be sequenced without the need for prior culturing. It is feasible to find genes responsible for the manufacture of novel chemicals that are inert in laboratory circumstances, such as those found in Streptomyces, which create secondary metabolites important to the pharmaceutical sector. As a result, it is easier to find new antibiotics or metabolites produced by microbes that could be useful in human or veterinary treatment (Challis 2014).

7.13.5 Vaccine Manufacturing

NGS is helping to untangle the intricate processes of viral evolution and host responses to these viruses, increasing the chances of a successful vaccine development. Several veterinary vaccines have been reported to be contaminated with adventitious agents, including feline endogenous retrovirus in feline- and canine-attenuated vaccines (Yoshikawa et al. 2014) and bovine viral diarrhea virus in swine fever vaccines (Bumbarov et al. 2016). Finally, additional challenges will emerge as NGS progresses. Single-molecule third-generation sequencing, for example, will likely remove the restriction of short reads, but it will also create new technical problems and design challenges. The introduction of next-generation sequencing (NGS) was a watershed moment in molecular biology and animal genetics.

7.13.6 Infectious Disease Diagnosis

Pathogen identification type using NGS is a highly sensitive approach. During a disease outbreak study in commercial chicken, we sequenced RNA and DNA extracted from pharyngeal mucosa using the ion torrent platform. Many challenges remain, including automation, standardizing technical protocols and bioinformatics pipelines, improving reference databases, and lowering cost and turnaround time, all of which would be required for widespread adoption of NGS in clinical microbiology laboratories. Next-generation sequencing opens up new possibilities for addressing issues related to pathogenic diseases. Apart from the genome sequences of the causal agents, we were able to identify the genomes of some live chicken vaccines, many bacteriophages, and fish and maize viruses. From better identifying the types of pathogens present in order to personalize therapies to understanding the

genetic basis of parasite resistance and devising new therapeutic approaches, next-generation sequencing technology opens up new possibilities.

7.13.7 Prediction of Diseases Outbreaks

For instance, NGS and bioinformatics approaches have been used to identify outbreak origins. NGS has also made it possible to detect copy number variations (CNV), which have opened up new possibilities for studying genes linked to complex characteristics in livestock. Whole-genome resequencing data identified 14,821 CNVs and 487 differential CNVRs in Holstein bulls with extraordinarily high and low estimated breeding values (EBVs) for milk protein percentage and fat percentage.

7.13.8 Study of Bovine and Other Ruminant Gut Microbiota

Different NGS platforms were utilized to assess the rumen microbial community in cows. While most of the studies were related to the rumen bacterial diversity (Jami et al. 2013; Zhou et al. 2017), very few of them studied the composition of rumen fungal community. Identification and characterization of rumen microbiota will help in developing better nutritional systems. Animal by-product meals from the rendering industry could be the most promising and suitable alternative to FM ingredients in aquaculture practice due to their high content of essential amino acids and water-soluble proteins.

7.13.9 Study of Animal Transcriptome

RNA sequencing is performed on such platforms as Roche454 or Illumina. In 2007, Roche was the company which developed the first RNAseq technique based on pyrosequencing. Currently, the most commonly used RNAseq technique is the chemical synthesis used by Illumina. In 2015, this technique provided 90% of all RNA sequences worldwide. The knowledge of mRNA or protein structures associated with meat quality facilitates the construction of practical biomarkers to assess and predict individual traits. Research in the field of poultry molecular biology is also an example of improving the efficacy of protecting animals against pathogens. Similar analyses are conducted in other animal species.

7.13.10 Exotic Diseases Identification

To prevent disease from spreading through the movement of animals and animal products,

Lumpy skin disease (LSD) is currently regarded as one of the most serious hazards to the Indian livestock industry. Because of its speed, sensitivity, and general nature, NGS offers a distinct advantage over other traditional pathogen detection approaches. Preexport screening for African swine fever, porcine brucellosis, classical swine fever, porcine respiratory and reproductive syndrome, porcine epidemic diarrhea virus, porcine delta corona virus, transmissible gastroenteritis, swine vesicular disease, swine influenza A virus, Teschovirus encephalomyelitis (TVE), and Aujeszky's disease is required when importing live (DAHD 2017).

7.13.11 Endangered Species Genomics

The goal of NGS is to sequence species that are on the verge of extinction, as well as extinct organisms, and human predecessors. Elephants are found in three species: two African and one Indian. The sequence of the elephant endotheliotropic herpesvirus (EEHV), which causes the lethal hemorrhagic sickness in elephants, was one of the targets of elephant genome study. More rigorous analyses of population demographic history and adaptive variation related with fitness and local adaptation are possible thanks to next-generation sequencing (NGS) and the collection of genome-wide data (Steiner et al. 2013).

References

Adessi C, Matton G, Ayala G et al (2000) Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. Nucleic Acids Res 28:E87. https://doi.org/10.1093/nar/28.20.e87

Alföldi J, Lindblad-Toh K (2013) Comparative genomics as a tool to understand evolution and disease. Genome Res 23(7):1063–1068. https://doi.org/10.1101/gr.157503.113

Alvarez-Garcia I, Miska EA (2005) MicroRNA functions in animal development and human disease. Development 21:4653–4662. https://doi.org/10.1242/dev.02073

Berry DP, Bermingham ML, Good M et al (2011) Genetics of animal health and disease in cattle. Ir Vet J 64:5. https://doi.org/10.1186/2046-0481-64-5

Boehmer JL, De Grasse JA, McFarl MA, Tall EA, Shefcheck KJ, Ward JL, Bannerman DD (2010) The proteomic advantage: label-free quantification of proteins expressed in bovine milk during experimentally induced coli form mastitis. Vet Immunol Immunopathol 138:252–266. https://doi.org/10.1016/j.vetimm.2010.10.004

Boichard D, Chung H et al (2012) Design of a bovine low-density SNP array optimized for imputation. PLoS One 7(3):e34130. https://doi.org/10.1371/journal.pone.0034130

Brew F (2014) Axiom Buffalo genotyping arrays, best for your genotyping program-now and tomorrow, vol 12. Plant and Animal Genome XXII Conference, Plant and Animal Genome. https://doi.org/10.1371/journal.pone.0185220

Bumbarov V, Golender N, Erster O, Khinich Y (2016) Detection and isolation of bluetongue virus from commercial vaccine batches. Vaccine 34:3317–3323. https://doi.org/10.1016/j.vaccine.2016.03.097

Cai W, Li C, Liu S, Zhou C, Yin H, Song J, Zhang Q, Zhang S (2018) Genome wide identification of novel long non-coding RNAs and their potential associations with milk proteins in Chinese Holstein cows. Front Genet 9:281. https://doi.org/10.3389/fgene.2018.00281

Cao C, Fan R, Zhao J, Zhao X, Yang J, Zhang Z, Xu S (2017) Impact of exudative diathesis induced by selenium deficiency on LncRNAs and their roles in the oxidative reduction process in broiler chick veins. Oncotarget 8:20695–20705. https://doi.org/10.18632/oncotarget.14971

Challis GL (2014) Exploitation of the Streptomyces coelicolor A3(2) genome sequence for discovery of new natural products and biosynthetic pathways. J Ind Microbiol Biotechnol 41(2):219–232. https://doi.org/10.1007/s10295-013-1383-2

Chen RJ, Kelly G, Sengupta A, Heydendael W, Nicholas B, Beltrami S et al (2015) MicroRNAs as biomarkers of resilience or vulnerability to stress. Neuroscience 305:36–48. https://doi.org/10.1016/j.neuroscience.2015.07.045

DAHD (2017) Veterinary certificate for import of live swine for breeding purposes into India in, D.F. Department of Animal Husbandry, ed.

Danesh J, Pepys MB (2009) C-reactive protein and coronary disease is there a causal link? Circulation 120(21):2036–2039. https://doi.org/10.1210/jc.2010-0473

Dekkers JCM (2012) Application of genomics tools to animal breeding. Curr Genomics 13:207–212. https://doi.org/10.2174/138920212800543057

Dratwa-Chałupnik A, Wojdyła K, Ożgo M, Lepczyński A, Michałek K, Herosimczyk A, Rogowska-Wrzesińska A (2020) Urinary proteome of newborn calves—new potential in non-invasive neonatal diagnostic. Animals 10:1257. https://doi.org/10.3390/ani10081257

Edea Z, Dadi H, Kim SW, Dessie T, Lee T, Kim H, Kim JJ, Kim KS (2013) Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers. Front Genet 4:35. https://doi.org/10.3389/fgene.2013.00035

Eisen MB, Spellman PT et al (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci 95(25):14863–14868. https://doi.org/10.1073/pnas.95.25.14863

El-Metwally S, Ouda OM, Helmy M (2014) Next-generation sequencing technologies and challenges in sequence assembly. Springer

Elsik CG et al (2009) The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science 324(5926):522–528. https://doi.org/10.1126/science.1169588

Escribano D, Horvatić A, Contreras-Aguilar MD, Guillemin N, Cerón JJ, Lopez-Arjona M, Hevia ML, Eckersall PD, Manteca X, Mrljak V (2019) Identification of possible new salivary biomarkers of stress in sheep using a high-resolution quantitative proteomic technique. Res Vet Sci 124:338–345. https://doi.org/10.1016/j.rvsc.2019.04.012

Frohlich T, Kirschbaum T, Thoenes U, Furrer F, Dietrich-Veenstra U, Seller M (2004) The LightTyper instrument: high-throughput genotyping of single-nucleotide polymorphisms. Biochemica Mannheim 2:9–11

Hanotte O, Jianlin H (2005) Genetic characterisation of livestock populations and its use in conservation decision-making. In: Ruane J, Sonnino A (eds) The role of biotechnology in exploring and protecting agricultural genetic resources, pp 89–96

Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I, Causey M, Colonell J, Dimeo J, Efcavitch JW, Giladi E, Gill J, Healy J, Jarosz M, Lapen D, Moulton K, Quake SR, Steinmann K, Thayer E, Tyurina A, Ward R, Weiss H, Xie Z (2008) Single-molecule DNA sequencing of a viral genome. Science 320(5872):106–109. https://doi.org/10.1126/science.1150427

Head SR, Komori HK, LaMere SA et al (2014) (2014) library construction for next-generation sequencing: overviews and challenges. Biotech 56:6177. https://doi.org/10.2144/000114133
Hillier LW et al (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432(7018):695–716. https://doi.org/10.1038/nature03154
Ho A, Murphy M, Wilson S, Atlas SR, Edwards JS (2011) Sequencing by ligation variation with endonuclease V digestion and deoxyinosine-containing query oligonucleotides. BMC Genomics 12:598. https://doi.org/10.1186/1471-2164-12-598
Holland PM, Abramson RD, Watson R, Gelfand DH (1991) Detection of specific polymerase chain reaction product by utilizing the 59-39 exonuclease activity of Thermus aquaticus DNA polymerase. Proc Natl Acad Sci U S A 88:7276–7280. https://doi.org/10.1073/pnas.88.16.7276
Holley RW, Apgar J, Everett GA et al (1965) Structure of a ribonucleic acid. Science 147:1462–1465. PMID: 5898068. https://doi.org/10.1126/science.147.3664.1462
Hu H, Fang Z, Mu T, Wang Z, Ma Y, Ma Y (2021) Application of metabolomics in diagnosis of cow mastitis: a review. Front Vet Sci 8:747519. https://doi.org/10.3389/fvets.2021.747519
Huang W, Long N, Khatib H (2012) Genome-wide identification and initial characterization of bovine long non-coding RNAs from EST data. Anim Genet 43:674–682. https://doi.org/10.1111/j.1365-2052.2012.02325.x
Hwang KC, Song KD, Kim TH, Jeong DK, Sohn SH, Lilleo HS, Han JY (2001) Genetic linkage mapping of RAPD markers segregating in Korean ogol chicken- white leghorn backcross population. Asian – Aust. J Anim Sci 14(3):302–306. https://doi.org/10.3233/JCB-15033
Ioannidis J, Donadeu FX (2018) Comprehensive analysis of blood cells and plasma identifies tissue-specific miRNAs as potential novel circulating biomarkers in cattle. BMC Genomics 19:243. https://doi.org/10.1186/s12864-018-4646-5
Jain M, Hugh EO, Paten B, Akeson M (2016) The oxford nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol 17:239. https://doi.org/10.1186/s13059-016-1103-0
Jami E, Israel A, Kotser A, Mizrahi I (2013) Exploring the bovine rumen bacterial community from birth to adulthood. ISME J 7:1069–1079. https://doi.org/10.1038/ismej.2013.2
Kemp SJ, Teale AJ (1994) Randomly primed PCR amplification of pooled DNA reveals polymorphism in a ruminant repetitive DNA sequence which differentiates Bos indicus and B. taurus. Anim Genet 25:83–88. https://doi.org/10.1111/j.1365-2052.1994.tb00084.x
Kern C, Wang Y, Chitwood J, Korf I, Delany M, Cheng H, Medrano JF, Van Eenennaam AL, Ernst C, Ross P, Zhou H (2018) Genome-wide identification of tissue-specific long non-coding RNA in three farm animal species. BMC Genomics 19:684. https://doi.org/10.1186/s12864-018-5037-7
Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res D1:D68–D73. https://doi.org/10.1093/nar/gkt1181
Kranis A, Gheyas AA, Boschiero C et al (2013) Development of a high-density 600K SNP genotyping array for chicken. BMC Genomics 14:59. https://doi.org/10.1186/1471-2164-14-59
Kumar R, Ali SA, Singh SK et al (2021) Peptide profiling in cow urine reveals molecular signature of physiology-driven pathways and in-silico predicted bioactive properties. Sci Rep 11:12427. https://doi.org/10.1038/s41598-021-91684-4
Lamers RJ, van Nesselrooij JH, Kraus VB et al (2005) Identification of a urinary profile associated with osteoarthritis. Osteoarthr Cartil 13:762–768. https://doi.org/10.1016/j.joca.2005.04.005
Laurent GS, Wahlestedt C, Kapranov P (2015) The landscape of long noncoding RNA classification. Trends Genet 31:239–251. https://doi.org/10.1016/j.tig.2015.03.007
Leo VC, Morgan NV, Bern D et al (2015) Use of next-generation sequencing and candidate gene analysis to identify underlying defects in patients with inherited platelet function disorders. J Thromb Haemost 13:643–650. https://doi.org/10.1111/jth.12836
Li Z, Ouyang H, Zheng M, Cai B, Han P, Abdalla BA, Nie Q, Zhang X (2016) Integrated analysis of long non-coding RNAs (lncRNAs) and mRNA expression profiles reveals the potential role

of lncRNAs in skeletal muscle development of the chicken. Front Physiol 7:687. https://doi.org/10.3389/fphys.2016.00687
Lim HJ, Kim HJ, Lee JH, Lim DH, Son JK, Kim ET, Jang G, Kim DH (2021) Identification of plasma miRNA biomarkers for pregnancy detection in dairy cattle. J Anim Reproduct Biotechnol 36:35–44. https://doi.org/10.12750/JARB.36.1.35
Livak KJ, Marmaro J, Todd JA (1995) Towards fully automated genome-wide polymorphism screening. Nat Genet 9:341–342. https://doi.org/10.1038/ng0495-341
Lu H, Giordano F, Ning Z (2016) Oxford nanopore MinION sequencing and genome assembly. Genom Proteom Bioinform 14:265–279. https://doi.org/10.1016/j.gpb.2016.05.004
Lyamichev V, Mast AL, Hall JG (1999) Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. Nat Biotechnol 17:292–296. https://doi.org/10.1038/7044
Ma L, Bajic VB, Zhang Z (2013) On the classification of long non-coding RNAs. RNA Biol 10:924–933. https://doi.org/10.4161/rna.24604
Marchi DM, Dalvit C, Targhetta C, Cassandro M (2006) Assessing genetic diversity in indigenous Veneto chicken breeds using AFLP markers. Anim Genet 37(2):101–105. https://doi.org/10.1111/j.1365-2052.2005.01390.x
Mardis ER (2008) Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9:387–402. https://doi.org/10.1146/annurev.genom.9.081307.164359
Marin JC, Saucedo CE, Corti P, Gonzalez BA (2009) Application of DNA forensic techniques for identifying poached guanacos (Lama guanicoe) in Chilean Patagonia. J Forensic Sci 54:1073–1076. https://doi.org/10.1111/j.1556-4029.2009.01087.x
Masoudi-Nejad A, Narimani Z, Hosseinkhan N (2013) Next-generation sequencing and sequence assembly. Springer, Methodologies and algorithms
Maxam AM, Gilbert W (1977) A new method for sequencing DNA. Proc Natl Acad Sci U S A 74(2):560–564. https://doi.org/10.1073/pnas.74.2.560
Mishra A, Verma M (2010) Cancer biomarkers: are we ready for the prime time? Cancer 2:190–208. https://doi.org/10.3390/cancers2010190
Mukhopadhyay C, Kumar D (2013) SNP chip development and genome wide association studies in livestock. In: Phenomic and genomic tools for analysis of livestock genome, vol 66
Muret K, Klopp C, Wucher V, Esquerré D, Legeai F, Lecerf F, Desert C, Boutin MS, Jehl F, Acloque H, Giuffra E, Djebali S, Foissac S, Derrien T, Lagarrigue S (2017) Long noncoding RNA repertoire in chicken liver and adipose tissue. Genet Sel Evol 49:6. https://doi.org/10.1186/s12711-016-0275-0
Naylor S (2003) Biomarkers: current perspectives and future prospects. Expert Rev Mol Diagn 3:525–529. https://doi.org/10.1586/14737159.3.5.525
Nicolazzi EL, Picciolini M, Strozzi F et al (2014) SNPchiMp: a database to disentangle the SNPchip jungle in bovine livestock. BMC Genomics 15:123. https://doi.org/10.1186/1471-2164-15-123
Nie L, Wu HJ, Hsu JM, Chang SS, Labaff AM, Li CW, Wang Y, Hsu JL, Hung MC (2012) Long non-coding RNAs: versatile master regulators of gene expression and crucial players in cancer. Am J Transl Res 4:127–150
Nothnick WB (2012) The role of micro-RNAs in the female reproductive tract. Reproduction 143:559–576. https://doi.org/10.1530/REP-11-0240
Orom UA, Nielsen FC, Lund AH (2008) MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. Mol Cell 30:460–471. https://doi.org/10.1016/j.molcel.2008.05.001
Ovilo C, Cervera MT, Castellanos C, MartõÂnez-Zapater JM (2000) Characterisation of Iberian pig genotypes using AFLP markers. Anim Genet 31:117–122. https://doi.org/10.1266/ggs.77.361
Pauley KM, Cha S, Chan EK (2009) MicroRNA in autoimmunity and autoimmune diseases. J Autoimmun 3:189–194. https://doi.org/10.1016/j.jaut.2009.02.012
Pelizzola M, Ecker JR (2011) The DNA methylome. FEBS Lett 585:1994–2000. https://doi.org/10.1016/j.febslet.2010.10.061

Peng Y, Chang L, Wang Y, Wang R, Hu L, Zhao Z, Geng L, Liu Z, Gong Y, Li J, Li X, Zhang C (2018) Genome-wide differential expression of long noncoding RNAs and mRNAs in ovarian follicles of two different chicken breeds. Genomics 7543:30268–30274. https://doi.org/10.1016/j.ygeno.2018.09.012

Perera FP, Weinstein IB (2000) Molecular epidemiology: recent advances and future directions. Carcinogenesis 21:517–524. https://doi.org/10.1093/carcin/21.3.517

Pisitkun T, Johnstone R, Knepper MA (2006) Discovery of urinary biomarkers. Mol Cell Proteomics 5:1760–1771. https://doi.org/10.1074/mcp.r600004-mcp200

Place RF, Li LC, Pookot D, Noonan EJ, Dahiya R (2008) MicroRNA-373 induces expression of genes with complementary promoter sequences. Proc Natl Acad Sci U S A 105:1608–1613. https://doi.org/10.1073/pnas.0707594105

Rabbini B, Tekin M, Mahdieh N (2014) The promise of whole-exome sequencing in medical genetics. J Hum Genet 59:5–15. https://doi.org/10.1038/jhg.2013.114

Rachagani S, Gupta ID, Gupta N et al (2006) Genotyping of β-Lactoglobulin gene by PCR-RFLP in Sahiwal and Tharparkar cattle breeds. BMC Genet 7:31. https://doi.org/10.1186/1471-2156-7-31

Ramos AM, Crooijmans RP et al (2009) Design of a high-density SNP genotyping assay in the pig using SNPs identified and characterized by next-generation sequencing technology. PLoS One 4(8):e6524. https://doi.org/10.1371/journal.pone.0086227

Ran M, Chen B, Li Z, Wu M, Liu X, He C, Zhang S, Li Z (2016) Systematic identification of long noncoding RNAs in immature and mature porcine testes. Biol Reprod 94:77. https://doi.org/10.1095/biolreprod.115.136911

Reverter A, Chan EK (2008) Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. Bioinformatics 24(21):2491–2497. https://doi.org/10.1093/bioinformatics/btn482

Roncada P, Piras C, Soggiu A, Turk R, Urbani A, Bonizzi L (2012) Farm animal milk proteomics. J Proteome 75:4259–4274. https://doi.org/10.1016/j.jprot.2012.05.028

Rottiers V, Näär AM (2012) MicroRNAs in metabolism and metabolic disorders. Nat Rev Mol Cell Biol 4:239–250. https://doi.org/10.1038/nrm3313

Rubio CP, Contreras-Aguilar MD, Quiles A, López-Arjona M, Cerón JJ, Martínez-Subiela S et al (2019) Biomarkers of oxidative stress in saliva of sheep: analytical performance and changes after an experimentally induced stress. Res Vet Sci 123:71–76. https://doi.org/10.1016/j.rvsc.2018.12.015

Rubio CP, Escribano D, Mainau E et al (2021) Changes in salivary biomarkers of oxidative status in calves at weaning and grouping. BMC Vet Res 17:373. https://doi.org/10.1186/s12917-021-03087-2

Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 94:441–448. https://doi.org/10.1016/0022-2836(75)90213-2

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74:5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Seidel G (2009) Brief introduction to whole-genome selection in cattle using single-nucleotide polymorphisms. Reprod Fertil Dev 22(1):138–144. https://doi.org/10.1071/rd09220

Shamsi FA, Chen Z, Liang J, Li K, Al-Rajhi A, Chaudhry I, Li M, Wu K (2011) Analysis and comparison of proteomic profiles of tear fluid from human, cow, sheep, and camel eyes. Invest Ophthalmol Vis Sci 52:9156–9165. https://doi.org/10.1167/iovs.11-8301

Shaughnessy RG, Farrell D, Stojkovic B et al (2020) Identification of microRNAs in bovine faeces and their potential as biomarkers of Johne's disease. Sci Rep 10:5908. https://doi.org/10.1038/s41598-020-62843-w

Shen L, Zhang S, Li Q, Fu Y, Tang G, Jiang Y, Li M, Wang J, Li X, Che L, Zhu L (2018) The landscape of non-coding RNA in an adult pig model of intrauterine growth restriction. Cell Physiol Biochem 50:1764–1778. https://doi.org/10.1159/000494794

Srikok S, Patchanee P, Boonyayatra S, Chuammitri P (2020) Potential role of MicroRNA as a diagnostic tool in the detection of bovine mastitis. Prev Vet Med 182:105101. https://doi.org/10.1016/j.prevetmed.2020.105101

Steiner CC, Putnam AS, Hoeck PE, Ryder OA (2013) Conservation genomics of threatened animal species. Annu Rev Anim Biosci 1:261–281. https://doi.org/10.1146/annurev-animal-031412-103636

Sun HZ, Wang DM, Wang B, Wang JK, Liu HY, Guan LL, Liu JX (2015) Metabolomics of four biofluids from dairy cows: potential biomarkers for Milk production and quality. J Proteome Res 14:1287–1298. https://doi.org/10.1021/pr501305g

Sundekilde UK, Frederiksen PD, Clausen MR, Larsen LB, Bertram HC (2011) Relationship between the metabolite profile and technological properties of bovine milk from two dairy breeds elucidated by NMR-based metabolomics. J Agric Food Chem 59:7360–7367. https://doi.org/10.1021/jf202057x

Talle SB, Yabuga WS, Fimland E, Syrstad O, Meuwissen T, Klungland H (2005) Use of DNA technologies for the conservation of animal genetic resources: a review. Acta Agr Scand A-An 55:1–8. https://doi.org/10.1080/09064700510009315

Te Pas MFW, Keuning E, Hulsegge B, Hoving-Bolink AH, Evans G, Mulder HA (2010) Longissimus muscle transcriptome profiles related to carcass and meat quality traits in fresh meat Pietrain carcasses. J Anim Sci 88(12):4044–4055. https://doi.org/10.2527/jas.2010-2952

TeleviCius M, JuozaitieneV MD, Antanaitis R, Rutkauskas A, Urbutis M, Baumgartner W (2021) Inline Milk lactose concentration as biomarker of the health status and reproductive success in dairy cows. Agriculture 11:38. https://doi.org/10.3390/agriculture11010038

Tian H, Zheng N, Wang W, Cheng J, Li S, Zhang Y, Wang JQ (2015) Identification of diagnostic biomarkers and metabolic pathway shifts of heat-stressed lactating dairy cows. J Proteome 125:17–28. https://doi.org/10.1016/j.jprot.2015.04.014

Tian H, Zheng N, Wang W, Cheng J, Li S, Zhang Y, Wang JQ (2016) Integrated metabolomics study of the Milk of heat-stressed lactating dairy cows. Sci Rep 6:24208. https://doi.org/10.1038/srep24208

Tobe VO, Taylor SL, Nickerson DA (1996) Single-well genotyping of diallelic sequence variations by a two-color ELISA-based oligonucleotide ligation assay. Nucleic Acids Res 24:3728–3732. https://doi.org/10.1093/nar/24.19.3728

Tong C, Chen Q, Zhao L, Ma J, Ibeagha-Awemu EM, Zhao X (2017) Identification and characterization of long intergenic noncoding RNAs in bovine mammary glands. BMC Genomics 18:468. https://doi.org/10.1186/s12864-017-3858-4

Tosser-Klopp G, Bardou P et al (2014) Design and characterization of a 52K SNP Chip for goats. PLoS One 9(1):e86227. https://doi.org/10.1371/journal.pone.0152632

Velho ALC, Menezes E, Dinh T, Kaya A, Topper E, Moura AA, Memili E (2018) Metabolomic markers of fertility in bull seminal plasma. PLoS One 10:13. https://doi.org/10.1371/journal.pone.0195279

Vezzi F (2012) Next-generation sequencing revolution challenges: search, assemble, and validate genomes. Ph.D., Universita degli Studi di Udine, Italy. https://doi.org/10.1007/978-981-15-0156-2_8

Voelkerding KV, Dames SA, Durtschi JD (2009) Next-generation sequencing: from basic research to diagnostics. Clin Chem 55:641–658. https://doi.org/10.1373/clinchem.2008.112789

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63. https://doi.org/10.1038/nrg2484

Wang X, Yang C, Guo F, Zhang Y, Ju Z, Jiang Q, Zhao X, Liu Y, Zhao H, Wang J, Sun Y, Wang C, Zhu H, Huang J (2019) Integrated analysis of mRNAs and long noncoding RNAs in the semen from Holstein bulls with high and low sperm motility. Sci Rep 9:2092. https://doi.org/10.18632/aging.103214

Weng B, Ran M, Chen B, He C, Dong L, Peng F (2017) Genome-wide analysis of long non-coding RNAs and their role in postnatal porcine testis development. Genomics 109:446–456. https://doi.org/10.1016/j.ygeno.2017.07.001

Whitfield PD, Noble PJM, Major H et al (2005) Metabolomics as a diagnostic tool for hepatology: validation in a naturally occurring canine model. Metabolomics 1:215–225. https://doi.org/10.1007/s11306-005-0001-3

Wilson KJ, De la Vega E (2005) The potential of microarrays to assist shrimp breeding and production: a review. Aust J Exp Agric 45:901–911. https://doi.org/10.1071/EA05060

Wu L, Fan J, Belasco JG (2006) MicroRNAs direct rapid deadenylation of mRNA. Proc Natl Acad Sci U S A 103:4034–4039. https://doi.org/10.1073/pnas.0510928103

Xu E, Zhang L, Yang H, Shen L, Feng Y, Ren M, Xiao Y (2019) Transcriptome profiling of the liver among the prenatal and postnatal stages in chickens. Poult Sci 98:7030–7040. https://doi.org/10.3382/ps/pez434

Yang W, Kang X, Yang Q, Lin Y, Fang M (2013) Review on the development of genotyping methods for assessing farm animal diversity. J Anim Sci Bio 4(2):1–6. https://doi.org/10.1186/2049-1891-4-2

Yeo JS, Lee JS, Lee CH, Jung YJ, Nam DH (2000) Identification of genetic markers for Korean native cattle (Hanwoo) by RAPD analysis, vol 5. Eunpyung-ku, Seoul, Korea, p 23

Yin Z, Lian L, Zhu F, Zhang ZH, Hincke M, Yang N, Hou ZC (2020) The transcriptome landscapes of ovary and three oviduct segments during chicken (Gallus gallus) egg formation. Genomics 112:243–251. https://doi.org/10.1016/j.ygeno.2019.02.003

Yoshikawa R, Shimode S, Sakaguchi S, Miyazawa T (2014) Contamination of live attenuated vaccines with an infectious feline endogenous retrovirus (RD-114 virus). Arch Virol 159(3):399–404. https://doi.org/10.1007/s00705-013-1809-1

You Z, Zhang Q, Liu C, Song J, Yang N, Lian L (2019) Integrated analysis of lncRNA and mRNA repertoires in Marek's disease infected spleens identifies genes relevant to resistance. BMC Genomics 20:245. https://doi.org/10.1186/s12864-019-5625-1

Yu L, Tai L, Zhang L, Chu L, Li Y, Zhou L (2017) Comparative analyses of long non-coding RNA in lean and obese pigs. Oncotarget 8:41440–41450. https://doi.org/10.18632/oncotarget.18269

Yue S, Ding S, Zhou J, Yang C, Hu X, Zhao X, Wang Z, Wang L, Peng Q, Xue B (2020) Metabolomics approach explore diagnostic biomarkers and metabolic changes in heat-stressed dairy cows. Animals 10:1741. https://doi.org/10.3390/ani10101741

Zeng B, Chen T, Xie MY, Luo JY, He JJ, Xi QY, Sun JJ, Zhang YL (2019) Exploration of long noncoding RNA in bovine milk exosomes and their stability during digestion in vitro. J Dairy Sci 102:6726–6737. https://doi.org/10.3168/jds.2019-16257

Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 4:Article17. https://doi.org/10.2202/1544-6115.1128

Zhang A, Sun H, Wang P, Han Y, Wang X (2012) Recent and potential developments of biofluid analyses in metabolomics. J Proteome 75:1079–1088. https://doi.org/10.1016/j.jprot.2011.10.027

Zhang G, Rupasri M, David W, Burim A (2021) A multi-platform metabolomics approach identifies urinary metabolite signatures that differentiate Ketotic from healthy dairy cows. Front Vet Sci 8:595983. https://doi.org/10.3389/fvets.2021.595983

Zhao Y, Li H, Fang S, Kang Y, Hao Y, Li Z, Bu D, Sun N, Zhang MQ, Chen R (2016) NONCODE 2016: An informative and valuable data source of long non-coding RNAs. Nucleic Acids Res 44(D1):D203–D208. https://doi.org/10.1093/nar/gkv1252

Zhou Z, Fang L, Meng Q, Li S, Chai S, Liu S, Schonewille JT (2017) Assessment of ruminal bacterial and archaeal community structure in yak (Bos grunniens). Front Microbiol 8:179. https://doi.org/10.3389/fmicb.2017.00179

Part II
Microbial Genomics and Novel Biotechnological Approaches to Enhance Animal Health

Chapter 8
Genomics Innovations and Advanced Technologies

Sunny Dholpuria, Santosh Anand, Kumar Siddharth Singh, Shikha Chaudhary, and Harsh Panwar

Abstract Genome technologies have generated an unparalleled quantity of resources to contribute to innovation, such as environmental sustainability, food security, alternative energy sources, and human health. It allows us to identify and categorize genetic heritage, responsiveness to environmental factors, and prospective therapeutic approaches. In this chapter, we discuss how advanced genomics can help us in understanding microbial diversity, antibiotic resistance, and the economics revolving around it. The chapter also provides an insight into miRNAs, siRNAs, and all other RNAs, which are relevant and useful in either diagnosis or therapy of disease in animals/humans. Further, the increasing importance of genome editing technologies like clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated protein 9 (Cas9), TALENs, ZFNs, etc., have been covered in the chapter, along with the advanced pathogen detection systems in dairy, functional foods involving animals, and advanced systems/technologies to monitor and improve quality, improve shelf life, detect pathogens, and ensuring traceability in food products.

S. Dholpuria
Department of Life Sciences, J. C. Bose University of Science and Technology, YMCA Faridabad, Faridabad, Haryana, India

S. Anand
Department of Dairy Microbiology, Phulo-Jhano Murmu College of Dairy Technology, Birsa Agricultural University, Dumka, Jharkhand, India

K. S. Singh (✉)
Institute for Microbiology, Leibniz University, Hannover, Germany

S. Chaudhary
Department of Bioinformatics, College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

H. Panwar
College of Dairy Science and Technology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India
e-mail: harshpanwar@gadvasu.in

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_8

Keywords Genomics · Microbial diversity · Genome editing tools · CRISPR-Cas system

8.1 Introduction

The recent technological advancements in genome sequencing have offered infinite number of opportunities for novel discoveries in basic and applied sciences. Technologies include novel molecular approaches, for example the use of miRNAs, siRNAs, and all other RNAs in diagnosis or therapy, personalized medicine that is attributable to several genome editing technologies, targeting genes of interest in both DNA and RNA profiling applications, clustered regularly interspaced short palindromic repeats also called as CRISPR/CRISPR-associated protein 9 (Cas9), transcription activator-like effector nucleases (TALENs), zinc-finger nucleases (ZFNs), and a wide range of sequencing platforms for short-read, high-throughput, and long-read single molecule and nanopore sequencing, etc. When properly implemented to examine changes in genetic makeup and genomic expression, these technologies provide the foundation for uncovering various pathways, marker genes, and mechanisms of several human diseases, as well as create the knowledge for precision medicine (Nallari 2020). Additionally, there are promising techniques such as sugar-based biochemical methods, biosensors, polymerase chain reaction (PCR) and based methods, molecular typing methods, etc. that are now a days being used in advanced pathogen detection systems, and being applicable in milk and milk products, probiotics associated with humans and animals.

8.2 Advanced Genomics Innovations in Understanding Microbial Diversity, Antibiotic Resistance

In the history of science, the breakthrough notion of cloning DNA directly from environmental samples to examine the complexity of natural microbial communities was pioneered for the first time over 30 years ago by Pace and colleagues (Pace et al. 1986) in 1986. At that time, the authors emphasized that, the makeup of the original microbial populations might be determined by analyzing whole or partial rRNA sequences. In terms of phylogeny and functioning, microbes are the most diverse group on Earth, occupying every imaginable niche. The great majority of these species are characterized using conserved marker genes such as small subunit ribosomal RNA or, more recently, shotgun sequencing (metagenomics) (Gilbert and Dupont 2011). When the scientists discovered metagenomics, they observed a new frontier in research emerge: the mining for novel chemical compounds from uncultured microorganisms, which accounts for more than 99% of microbial diversity. This novel notion in microbial science opened the scientific community's eyes to the astoundingly enormous repertoire of biochemical functions available in nature

that are yet to be found (Luana et al. 2018). Metagenomic projects have been integrated with next-generation sequencing (NGS) technology over the last 10 years and this has accelerated the advancement of microbial ecology (Tringe and Hugenholtz 2008). Current NGS methods provide multiple times higher throughput than traditional Sanger sequencing, and the technologies are rapidly increasing (Metzker 2010). This makes next-generation sequencing (NGS) one of the trendiest issues in biological sciences. Scientists can now unveil the mysteries of the lives of yet uncultured microbes with the help of the newly created discipline of metagenomics and high-throughput sequencing tools (Rajendhran and Gunasekaran. 2011).

With enormous increase in the antimicrobial resistance now a days, the bacteria are becoming resistant to commonly used antibiotics, such that even successful treatments are having the potential to fail and may compel us toward the frequent use of drugs of higher toxicity levels as last resort, such as colistin. The use of whole-genome sequencing for antibiotic susceptibility testing (WGS-AST) is now a potent alternative to culturing bacteria in culture to assess antimicrobial susceptibility profiles in clinical microbiology (Michelle et al. 2018). WGS-AST (whole-genome sequencing for antimicrobial susceptibility testing) has the capacity to predict every known resistance characteristic for a strain quickly, consistently, and accurately, while also providing rich monitoring data. It is also conceivable to perform AST after direct shotgun sequencing of clinical samples (metagenomic-AST), although it is complicated and costly (Bradley et al. 2015). Prior to sequencing, metagenomic-AST methods enhance a library of antibiotic resistance DNA fragments from complicated clinical samples (Doyle et al. 2018). Synthetic long-read technologies that are emerging nowadays depend on the barcoding of long DNA fragments to associate the resulting short reads during assembly (Kuleshov et al. 2016) may provide more precision, and cost savings, that make routine WGS-AST and routine metagenomic-AST possible.

8.3 The Economics of Next-Generation Sequencing and Emerging Markets

Next-generation sequencing (NGS) is a fast-evolving field of technology that has driven the research sector and has enormous therapeutic applications. It has transformed the ability to sequence nucleic acids by boosting the amount of sequence data that can be collected in a more cost-effective manner and in a lot less amount of time (Wetterstrand 2018). When compared to Sanger sequencing, next-generation sequencing has dramatically reduced time and cost while greatly increasing sequence yield. The method has transformed DNA sequencing to the point that a complete human genome may now be sequenced in 3 days for less than US $1000 (Kumar et al. 2019). NGS testing that includes multigene panels, and WGS are becoming more popular (Phillips et al. 2018a, b). However, just a few

surveys have looked into their economic worth (Schwarze et al. 2018). Although, several publications have recognized methodological issues in evaluating NGS testing, which may be a barrier to undertaking evaluations (Fugel et al. 2016). NGS tests can produce various results, as well as additional findings that might not be related to the primary reason for testing. Each of these discoveries may have diverse clinical trajectories, resulting in various economic consequences. Secondary results and variations of undetermined significance can have an effect on the cost and outcomes in a favorable or negative way as well. Furthermore, findings may have interaction consequences, resulting in a sum bigger than the parts (Phillips et al. 2018a, b). The vast data generated by sequencing programs demonstrate that NGS is advantageous to health care systems and can be used to construct cost-effectiveness models evaluating other NGS approaches such as WGS. The inclusion of genomics data to other linked datasets may aid in understanding costs and outcomes, allowing health economists to develop more precise financial projections in future (Wordsworth et al. 2018).

8.4 Host–Microbe Interactions and Role of Surface Proteins

Interactions between the animals and microbes are well-known with many microbes living inside their hosts. Also, both negative and positive interactions with specific microbes are well-documented. The significance of host–microbe interactions of individual microbes remains relevant and this has often been attributed to microbial proteins (Sengupta et al. 2013). The research into surface proteins and recent advances has improved our understanding of host–microbe interactions at the molecular level. Microbes inhabiting/interacting within the host use their surface proteins to modulate the host cell expression of genes, immunity, mucus barrier, etc. to establish themselves or as a part of their symbiotic/commensal relationship with the host. Many microbes are pathogens which have developed specific strategies to establish themselves inside the host and evade host immunity for their survival (Sen et al. 2016). Beneficial microbes such as probiotics are known to help maintain host health, improve immunity, and evade pathogens, during their residence inside the host. Both of these categories of microbes have been known to bear many surface proteins such as mucus-binding protein (Singh et al. 2017), fibronectin-binding protein (Bisht et al. 2018), surface-layer protein (Choudhary et al. 2023), etc., which help in their interaction with the host.

These surface proteins help in cross talk as molecular effectors, adhesion to the host mucosa, and modulating the host immunity. Many of these surface proteins are adhesins which facilitate adhesion to the host and are used both by the probiotic bacteria and the pathogens. Apart from nutrient limitation and production of antimicrobial compounds, probiotics are also able to combat the pathogens by competitive exclusion from the gastrointestinal tract of their host. Recently, it has also been shown that the surface adhesins from probiotics such as mucus-binding protein, fibronectin-binding protein, or even prebiotics, etc., can be used to effectively

exclude pathogens (Singh et al. 2018; Bisht et al. 2018; Anand et al. 2018). Some of these adhesins also have multifunctional roles, which are being slowly uncovered. Recently, the role of microbiome and its significance in normal health, diseases, and dysbiosis manifestations of human and animals has also gained attention. The microbiota helps maintain normal homeostasis of the host, but has also been shown to enhance host susceptibility to particular pathogens and thus contribute to disease manifestation. This emerging domain of research has benefited directly from the advances in genome sequencing, in silico annotations, recombinant DNA technology, protein–protein interaction studies, biophysical techniques, and cell-line-based models of disease or infection (Sen et al. 2016). The host–microbe interactions have also been known to be influenced by the host genetics such as secretor status (for a particular antigen on blood cells), the epigenetics, and the type of diet such as milk and milk components (Kumar et al. 2015; Kumbhare et al. 2017).

8.5 Milk Bioactive Components in Host–Microbe Interactions

Human milk is an essential diet and a rich source of nutrition and early age microbiome for infants. Apart from supplying proteins, fats, carbohydrates, and minerals, it has been recently shown to contain many bioactive components. These bioactive components such as human milk oligosaccharides (HMOs) and bioactive peptides have complex role in the growth, immunity, and development of the children. HMOs remain undigested in human host and act as prebiotics, thereby promoting the growth of probiotics and thus indirectly improving host health (Urashima and Taufik 2011). HMOs not only act as energy source for prebiotics, but also enhance the adhesion of probiotics to the host gut and modulate the host gut barrier function through changes in the tight junctions (Chichlowski et al. 2012).The similarity of HMOs with cell surface antigens and their role as “decoys” for these antigens has been implicated for this attribute and has also been shown to be an important factor in susceptibility to viral infections (Naarding et al. 2005). HMOs have also been known to play role in immunomodulation of the host by changing its pattern of gene expression (Kuntz et al. 2008).

Milk bioactive peptides are derived from proteolysis and fermentation of milk proteins and are broadly classified as preabsorptive bioactive peptides and postabsorptive bioactive peptides, depending upon their action before or after absorption in the host gut (Singh et al. 2021).These peptides have been found to have multiple health effects on the host, such as increase of growth of probiotics in host gut by acting as prebiotics (Liepke et al. 2002), protection of host gut mucosa from binding of intestinal pathogens (Nakajima et al. 2005), as antimicrobials (Pellegrini et al. 1999), and in improvement of intestinal calcium absorption as in case of caseinophosphopeptides (FitzGerald 1998). Milk-derived postabsorptive bioactive peptides have been found to act as Angiotensin I-converting enzyme

(ACE)-inhibitory peptides, which regulate hypertension and cardiovascular health (FitzGerald et al. 2004). Milk-derived peptides also comprise of opioid receptor ligands or opioid peptides, which are implicated to have calming effect, stress adaptation, gut development, electrolyte transport, etc. in the infants (Wada and Lönnerdal 2014). Some peptides such as Glycomacropeptide have been reported to have antithrombotic and anti-inflammatory effects (López-Posadas et al. 2010) and few milk-derived small peptides have immunostimulating and antioxidant property (Singh et al. 2021). Altogether, the bioactive properties of a universally accepted diet such as milk have profound effects on host–microbe interactions. Apart from these specific factors which affect the organism at the cellular/organ level, there is a growing realization of the overarching effect of environment and other organisms (such as animals) on host–microbe interactions in humans. One such outcome of such "cause-effect" phenomenon is antibiotic resistance which has become a global hazard and needs concerted efforts under an approach called as "One-Health".

8.6 One-Health Concept and Antimicrobial Resistance

"One-Health" concept refers to an interdisciplinary, collaborative, and multidomain approach for measures taken at local and global scale to mitigate an issue (One health 2008). Antimicrobial resistance (AMR) could be seen as a complex interplay of host–microbe-environment interactions and involves an exchange of microbes and antibiotic resistance determinants (ARDs) across ecological niches involving animals, humans, and the environment (Singh et al. 2021). ARDs refer to antibiotic resistance genes, integrons, etc. which form the physical basis for the transfer of antibiotic resistance across these niches. AMR kills at least 75,000 people every year across the world and puts extra burden on the health infrastructure and people's health and quality of life. AMR has many confounding factors, but indiscriminate usage of antibiotics has been the major reason for rapid emergence of AMR. Another related observation is the involvement of animal carriers and zoonotic sources during disease outbreaks (He et al. 2021). For example, many antibiotics used during raising livestock and metaphylaxis (antibiotics as growth enhancers) are reserved for treatment of terminally ill human patients. Since, the amount of antibiotics used in livestock is much higher and a lot of it ends up polluting our environment, the emergence of resistant microbes is expedited. This has been seen in case of many clinically important pathogens such as *E. coli, Salmonella spp.,* and *K. pneumoniae* which have gained resistance due to overuse of antibiotics and are often found in environmental samples, often identified by high-throughput technologies such as next-generation sequencing (Nordmann 2014; Martínez et al. 2015).

The interrelatedness of the complex landscape of AMR also involves stakeholders like bacteriophages which have also been implicated in aiding the transfer of ARDs. They have been found to mobilize the resistance genes through horizontal gene transfer (Brabban et al. 2005) and reported to transfer single/multidrug resistance traits to *E. coli* (Colomer-Lluch et al. 2011), *Salmonella enterica* (Schmieger

and Schicklmaier 1999), and *Pseudomonas aeruginosa* (Blahová et al. 2000). High gene mobilization traits and the presence of bacteriophages in complex ecological niches such as sewage, which also harbors a huge number of resistant microbes, further escalate their risk with regard to AMR (Singh et al. 2021). Sewage and manure are other niches that get a lot of inputs from humans or animals and harbor a huge number of resistant microbes and ARDs. They could be considered both as a source and sink/reservoir for many resistance genes and often come in contact with humans either through the air, irrigation water, etc., thereby increasing the load of resistant microbes in agriculture and directly to humans. Soil and aquatic bodies are also important reservoirs for resistant microbes because they receive inputs from manures, animal farms, irrigation water, pharmaceutical plants, humans, hospitals, and other antibiotic-contaminated niches through surface runoff (Rodgers et al. 2019). These become important, especially in the context of direct exposure to farmers during produce and transport of antibiotic/resistant microbe-contaminated agricultural produce. Clean water is an issue worldwide and thus, it poses a direct risk to interacting humans and animals. Some events such as mass gatherings for sports/religious activities/etc. also contribute to the rapid exchange of resistant microbes and have been found to cause a spike in the profile of resistant microbes such as *Enterobacter, Citrobacter, Serratia, and Klebsiella* (Jani et al. 2018). All of the aforementioned factors involve humans and the issue is further aggravated due to indiscriminate antibiotic prescription, poor hygiene, and other anthropogenic factors (Singh et al. 2021).

In these complex dynamics of AMR, which transcends different ecological niches, the "One-Health" approach is the most potent tool to deal with it. This concept, when applied to human stakeholders, is starkly visible in antibiotic stewardship, public awareness programs, and high-throughput surveillance technologies against AMR. This provides a dynamic picture of the endemic microbes bearing the burden of AMR in each niche and helps in designing the control and mitigation strategies (Critchley and Karlowsky 2004). This leads to improved general behavior toward antibiotics, prescription patterns, and the development of better waste disposal/recycling protocols, altogether helping toward robust public health. Other stakeholders are the regulatory agencies, which would benefit from the "One Health approach." This would be achieved by the integration and implementation of policies designed under the purview of responsible management of wastes, byproducts, and runoffs from antibiotic-contaminated farms, livestock facilities, drug production plants, hospitals, and sewage (Fisman and Laupland 2010). The bracket framework of "One Health" is an effective solution for the design and implementation of coordinated action against AMR, and the positive/negative role of host–microbe environment is efficiently managed.

8.6.1 Microbial ncRNAs

Noncoding RNAs (ncRNAs) are very well-reported in the eukaryotes, where they found in different forms like microRNAs, siRNAs, piRNAs, snoRNAs, snRNAs, and the long ncRNAs. If we consider the case of human, it has been estimated that around 98% of all transcriptional products come under the category of noncoding RNAs. All these play number of different roles in maintaining the complexity and homogeneity of different types of cells via like RNA interference, cosuppression, transgene silencing, imprinting, methylation, protein functioning, and many more (Mattick 2001). In comparison to eukaryotic system, the prokaryotic system is less complex and does not require much of regulation. In this context, it was thought that in the prokaryotic system, ncRNA is present in the form of housekeeping RNAs such as rRNA and tRNA, just like they are found in eukaryotes. But over the last two decades, a number of studies have revealed that a number of ncRNAs do get expressed in different bacterial systems. These microbial ncRNAs are found to exist in short as well as long forms and control various functions inside the bacterial cells. The primary difference between short and long forms is generally based on the fact that those RNAs are longer than 200 nucleotides and do not code for any proteins (Zampetaki et al. 2018). Short noncoding RNAs have been found to be associated with controlling a number of intracellular functions in microbes like in *Escherichia coli,* a small ncRNA 6S regulates RNA polymerase activity via interacting with sigma-70 (σ^{70}) subunit during the stationary growth phase of bacteria (Wassarman and Storz 2000). At the translation level also, role of small ncRNA can be seen in *Salmonella* and *Klebsiella* species; two ncRNAs, DsrA and RprA, have been found to positively regulate the RpoS (a stationary phase sigma factor) translation (Majdalani et al. 2001). A number of metabolic processes are also been controlled by ncRNA; like in the case of iron metabolism, a 90-nucleotide-long RNA RyhB has been found to downregulate a set of iron-storage and iron-using proteins when the iron is limited in the surroundings of bacteria. Using antisense approach, RyhB causes truncation of the operon *sdhCDAB* which encodes for succinate dehydrogenase along with other genes (Massé and Gottesman 2002). These small ncRNAs are found to be crucial during carbon starvation, like the expression of a small ncRNA, CrfA, in *Caulobacter crescentus* gets enhanced during carbon starvation which in turn regulates the expression of 27 other genes important for the survival of bacteria (Landt et al. 2010). In motility, they also play an equally important role as in the case of Salmonella Typhimurium, the expression of small ncRNA, ScoC, gets enhanced, thereby regulating the genes involved in flagellar synthesis (Fuentes et al. 2015). Another interesting example comes in the case of *Staphylococcus epidermidis* biofilm communities where a small ncRNA, RsaE, helps in formation of biofilm via releasing extracellular DNA as a supporting material (Schoenfelder et al. 2019).

Apart from these small ncRNAs, large ncRNAs are also showing their microbial world (Harris and Breaker 2018). A large ncRNA OLE (approx. 610 nucleotide) has been found to protect extremophiles from the toxicity caused by the presence of

alcohol (Wallace et al. 2012). Even after the identification of many other long ncRNA in different bacterial species, like GOLLD (giant, ornate, lake-, and Lactobacillales-derived) RNA and ROOL (rumen-originating, ornate, large) RNAs, their biological importance is still need to be explored (Harris and Breaker 2018).

8.6.2 ncRNAs and Microbial Pathogenicity

Microbial species are well-known for causing diseases in animals, plants, and humans. If we talk about humans only, we have a well-defined immune system to fight against these infectious agents. Whenever our immune system is compromised, we become more prone to fall ill. Even when our immune system is perfectly working, the infectious agents develop various strategies to overcome these barriers in order to survive and spread. In the same context, it seems like ncRNAs present in bacterial species are not only controlling their biological functions but they are also playing a role in causing various infections to animals as well as humans in uropathogenic *E. coli* strain UTI8, seven small noncoding RNAs have been identified for their putative role in providing type II persister formation which in turn help the bacteria to survive under lethal antibiotics and stress conditions (Zhang et al. 2018). *Mycobacterium tuberculosis*, which is one of the most successful human pathogens responsible in terms of causing more deaths than any other microbes, also shows the existence of small noncoding RNAs which confers pathogenicity to this bacteria (Arnvig and Young 2012). In has been seen that the expression of a small ncRNA, DrrS, in the *Mycobacterium tuberculosis* helps the bacteria to survive against oxidative and nitric oxide stresses in the intramacrophage environment (Salina et al. 2018). Expressions of two another ncRNAs, MTS0997 and MTS1338, are found to be crucial for the adaptation and virulence of *M. tuberculosis* (Shepelkova et al. 2021). Apart from these M. tuberculosis harbor, a battery of miRNA expression during its infection and these small RNAs help the bacteria in modulating critical pathways of the host cell autophagy mechanism (Wang et al. 2020). Bacterial communication depends on a process known as quorum sensing (QS) which involves production, secretion, and detection of signaling molecules and the role of ncRNA in regulating QS system has also been seen observed (Repoila and Darfeuille 2009). In case of *Vibrio cholera* (human pathogen responsible for food- and water-associated infections), the interaction between four ncRNAs named Qrr (quorum regulatory RNAs) and hapR mRNA found crucial for the pathogenicity of the bacteria (Repoila and Darfeuille 2009; Bardill et al. 2011). *Salmonella*, one of the leading causes for the gastrointestinal problem, also shows the presence of ncRNAs which determines the virulence nature of the bacteria like ncRNA, RaoN, contributes in overcoming the intracellular stress conditions of the macrophage (Lee et al. 2013). Another ncRNA, IsrM, located at *Salmonella* pathogenicity islands (commonly found in pathogenic strains), helps the bacterial invasion in epithelial cells as well as intracellular replication inside macrophages (Gong et al.

2011). Likewise, a number of ncRNAs have been reported in different pathogenic bacteria conferring virulence nature via modulating host response or controlling the expression of genes responsible for pathogenicity (González Plaza 2020; Wachter et al. 2020).

8.6.3 Combating Microbial Infection Via Targeting ncRNAs

In order to treat any bacterial infection (major or minor), the first thing that comes in mind is antibiotic and soon after the discovery of first true antibiotic Penicillin (in year 1928), more and more antibiotics have been discovered/manufactured. Most of these antibiotics generally target the enzymes/proteins involved in bacterial DNA replication, transcription, translation, and cell wall synthesis (O'Rourke et al. 2020). However, excessive use of these antibiotics over the years have led to serious issue of antimicrobial resistance and which in turn has also resulted in evolution of microbes in terms of pathogenicity (Abushaheen et al. 2020). New strategies are being explored to overcome this issue and recently identified bacteriophage therapy appears to be quite promising; however, certain regulatory issues are still major obstacle for the implementation of phage therapy (Chanishvili and Aminov 2019). With the advent of new techniques/technologies, more novel methods can be looked for aiming the ncRNAs as a strategy to combat bacterial infection. One such futuristic technology involves a target-based gene editing system known as CRISPR/Cas9 technology. Although, this system is a part of bacterial defense mechanism (Barrangou et al. 2007) but just like antibiotics (also being produced by bacteria), this system can be modulated to work against them. Being guided by small RNAs, Cas9 (a type of nuclease) helps in achieving target-specific gene editing and destruction (Tasan and Zhao 2017; Razzaq and Masood 2018). A similar approach can be used against ncRNAs of pathogenic bacteria to stop them for causing the infection and this will help in overcoming the problem of antibiotic resistance (Beisel et al. 2014). However, just like phage therapy, the application holds a tricky part which is designing a site-specific-guided RNA with utmost precision against the target sequence, and certain undesirable results have also raised concern on the use of this technology (Pineda et al. 2019). In the current scenario, CRISPR technology seems to be the future of antibacterials potentially capable of targeting any pathogenic bacteria (Greene 2018).

8.7 Advances in Detection of Pathogens in Dairy Products

Contamination of dairy products is a worldwide concern and usually due to environment, and human-related animal handling (Sofos 2008). Higher demands, global trade, climate change, and modified farm practices can direct a favorable situation for transmission of foodborne pathogens in between farm to fork chain due to a small

mistake within processing steps and supply of products globally, resulting endemic infectious disease outbreaks. Researchers have recognized various foodborne pathogens in dairy products evidencing a severe threat to human illness and death worldwide, having a significant economic impact (Fusco and Quero 2014). Hence, detection of these microbes is an essential task for accomplishment of food safety objective.

Ideal identification techniques should be quick, efficient, economical, highly sensitive, specific, and nonlabor intensive to qualify products before reaching to consumers. The conventional methods are usually the gold standard to work upon but they are laborious and need skilled technical hands. Modern techniques such as biosensors, immunological assays, and macromolecule-based (nucleic acid) methods are being developed and refined to overcome traditional methods' limitations (Sai-Anand et al. 2019). The microbiological detection techniques are classified into culture-dependent and culture-independent methods. Pathogens in milk and milk products can be detected either by a conventional culture-based approach involving enumeration, isolation, characterization, and identification or through a culture-independent approach comprising isolation, purification of nucleic acid, and molecular identification (Sudhakaran and Anand 2020).

Herein, a brief description of the most commonly used promising techniques and their application to milk and milk products is illustrated below.

8.7.1 Sugar-Based Biochemical Methods

In diagnosis, Vitek 2 Compact (bioMerieux), BD Phoenix (BD Diagnostics), Analytical Profile Index (API) 50 CHL (bioMerieux), and API 20E (bioMerieux) are some efficient kit methods for analyzing the identity of dairy microorganisms (Franco-Duarte et al. 2019).

8.7.2 Biosensors

Biosensors are devices that can transform the biological response of an incorporated biorecognition element such as antibodies, enzyme, nucleic acids, aptamers, bacteriophage, organelle, microorganism, etc. to a measurable signal by a physical transducer like optical, electrochemical, or mass-based, connected with an information output system Rubab et al. 2018). Electrochemical, optical, and mass-sensitive biosensors are the common biosensors used in dairy research (Perumal and Hashim 2014). Electrochemical biosensors containing amperometric or potentiometric transducers are quite inexpensive, unaffected by milk turbidity and used for pathogen detection in samples (Muniandy et al. 2019). Enzyme like alkaline phosphatase or horseradish peroxidase, upon oxidation of a substrate develops an electrical current that is monitored, and the signal strength is proportional to the microbial load

(Mortari and Lorenzelli 2014). Fluorescence and surface plasmon resonance (SPR)-based optical biosensors are also found to have excellent sensitivity (Rajapaksha et al. 2019). Commercially optical biosensors like BioFlash® (Rider et al. 2003), BIACORE Q, and Spreeta™, are adopted for the identification of foodborne pathogens (Mortari and Lorenzelli 2014). Mass-sensitive biosensors use piezoelectric quartz crystal microbalance (QCM) and wireless biosensing magnetoelastic transducers for testing milk microbes (Shen et al. 2011).

8.7.3 Polymerase Chain Reaction (PCR) and Based Methods

Polymerase chain reaction (PCR) is a molecular technique that amplifies a specific DNA sequence (nucleic acid, or cDNA generated from RNA as template) by using primers (universal, species-specific, or genus-specific) by mimicking the in vitro replication conditions in the presence of DNA polymerase (O'Sullivan et al. 2013). PCR is well-established method by International Organization for Standardization (ISO) for the detection of foodborne pathogens (ISO 2005a, b, 2006a, b, 2011a, b). In the dairy industry, qPCR and PCR techniques are commonly implemented due to their quick and cost-effective procedure and require basic skill to perform. Real-Time PCR can be used in both culture-dependent and -independent methods. The qPCR technique can be used to enumerate probiotics, can differentiate between dead and viable cells, and to study the associations between starter cultures in fermented milk products (Herbel et al. 2013). In a study, Qian et al. (2016) used droplet-based digital PCR (dPCR)to detect *Bifidobacterium* and *Lactobacillus* in the breast milk. Similarly, absolute quantification of viable lactic acid bacteria in fecal samples and soft cheese was detected by dPCR (Gobert et al. 2018). Foodborne pathogens like *E. coli* O157:H7, *Salmonella* (Jany and Barbier 2008), and *Bacillus cereus* in milk (Porcellato et al. 2016; Gobert et al. 2018) were detected by PCR. Polymerase chain reaction techniques can accurately symbolize physiological states of microorganisms and their viability during products processes (Randazzo et al. 2009; Matijasic et al. 2010; Bove et al. 2011). Food pathogens like *Bacillus cereus, Listeria monocytogenes*, and *Staphylococcus aureus* can be detected by multiplex qPCR with High-Resolution Melt (HRM) analysis in the milk sample (Forghani et al. 2015).

The viability PCR (vPCR) approach is another method used to compute the viable and dead cells. Here, intercalating agents like propidium monoazide (PMA) and ethidium monoazide (EMA) are used. Earlier, reported research showed good correlations with the plate counts and viability, especially during the enumeration of viable cells in probiotic products or fermented milk products using EMA-PCR and PMA-PCR (Meng et al. 2010; Fittipaldi et al. 2012; Sohier et al. 2014).

Other PCR-based methods like Temporal Temperature Gradient Gel Electrophoresis (TTGE) and Denaturing Gradient Gel Electrophoresis (DGGE) methods use sequence-specific separation of amplified 16S rDNA fragments (Naum and Lampel 2016). TTGE is used for the identification of microorganisms in commercial starter

cultures, fermented milk products, and types of microflora in variety of cheeses (Ogier et al. 2002). On the other hand, DGGE has been applied to analyze the microbial diversity of a product and to study fermentation ecology (Cocolin et al. 2013).

8.7.4 Molecular Typing Methods

Numerous molecular typing techniques have been employed for the identification and classification of bacteria up infraspecific level. Popularly known genetic-based molecular methods are DNA fingerprinting techniques like pulsed-field gel electrophoresis (PFGE), ribotyping, randomly amplified polymorphic DNA (RAPD), and amplified fragment length polymorphism (AFLP). These methods can detect of DNA polymorphisms between species or strains and differ in their dynamic range of taxonomic discriminatory power, reproducibility, ease of interpretation, and standardization. These methods are successfully used to identify LAB and bifidobacteria isolated from fermented food products as well as from the human gastrointestinal tract (McCartney 2002; Sharma et al. 2020). Conventional typing is based on phenotypes, such as serotype, biotype, phage type, bacteriocin typing, or antibiogram (Sabat et al. 2013). Later on, various methods like sodium dodecyl sulfate-poly acrylamide gel electrophoresis (SDS PAGE), multilocus enzyme electrophoresis (MLEE), mass spectrometry (MS), matrix-assisted laser desorption/ionization-time of-flight (MALDI-TOF) MS, Infrared or Raman spectroscopy, Fourier transform Infrared spectroscopy, etc. are employed, allowing simultaneous identification of dairy and probiotic bacteria containing multiple strains (Smith and Osborn 2009; Fusco and Quero 2014; Chifiriuc et al. 2017).

8.8 Conclusion

Advanced genomic innovations and technologies have been playing a significant role and have revolutionized various areas of science including the fields of medical and diagnostic relevance such as antibiotic resistance, and have significantly enhanced our understanding of microbial diversity. Apart from this, advances in technologies like genome sequencing, *in-silico* annotations, recombinant DNA technology, protein–protein interaction studies, biophysical techniques, etc. are greatly contributing to emerging fields of research nowadays. This is further on the bigger picture, contributing toward one health concept which eventually promises a healthier environment as a whole. These technologies are also helping us to move toward having targeted treatments/precision medicine at the genomic level. Additionally, there have been tremendous technological advancements such as biosensors, PCR-based methods, molecular typing methods, etc., that help us to detect pathogens in dairy products well in time, which can otherwise cause major

foodborne illnesses if they go undetected. Taken together, the contributions of these advances in research are enormous, and they have caused numerous breakthroughs in science and will continue to do so.

References

Abushaheen MA, Fatani AJ, Alosaimi M, Mansy W, George M, Acharya S, Rathod S, Divakar DD, Jhugroo C, Vellappally S, Khan AA (2020) Antimicrobial resistance, mechanisms and its clinical significance. Dis Mon 66:100971

Anand S, Mandal S, Singh KS, Patil P, Tomar SK (2018) Synbiotic combination of lactobacillus rhamnosus NCDC 298 and short chain fructooligosaccharides prevents enterotoxigenic *Escherichia coli* infection. LWT 98:329–334. https://doi.org/10.1016/j.lwt.2018.08.061

Arnvig K, Young D (2012) Non-coding RNA and its potential role in mycobacterium tuberculosis pathogenesis. RNA Biol 9:427–436

Bardill JP, Zhao X, Hammer BK (2011) The *Vibrio cholerae* quorum sensing response is mediated by Hfq-dependent sRNA/mRNA base pairing interactions. Mol Microbiol 80:1381–1394

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315:1709–1712

Beisel CL, Gomaa AA, Barrangou R (2014) A CRISPR design for next-generation antimicrobials. Genome Biol 15(11):1–4

Bisht S et al (2018) Expression of fibronectin-binding protein of L. acidophilus NCFM and in vitro refolding to adhesion capable native like protein from inclusion bodies. Protein Expr Purif 145: 7–13

Blahová J, Králiková K, Krcméry V, Jezek P (2000) Low-frequency transduction of imipenem resistance and high-frequency transduction of ceftazidime and aztreonam resistance by the bacteriophage AP-151 isolated from a Pseudomonas aeruginosa strain. J Chemother (Florence, Italy) 12(6):482–486. https://doi.org/10.1179/joc.2000.12.6.482

Bove CG, Lazzi C, Bernini V, Bottari B, Neviani E, Gatti M (2011) cDNA-amplified fragment length polymorphism to study the transcriptional responses of *lactobacillus rhamnosus* growing in cheese-like medium. J Appl Microbiol 111(4):855–864

Brabban A, Hite E, Callaway T (2005) Evolution of foodborne pathogens via temperate bacteriophage-mediated gene transfer. Foodborne Pathog Dis 2(4):287–303. https://doi.org/10.1089/fpd.2005.2.287

Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, Earle S, Pankhurst LJ, Anson L et al (2015) Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and mycobacterium tuberculosis. Nat Commun 6:10063

Chanishvili N, Aminov R (2019) Bacteriophage therapy: coping with the growing antibiotic resistance problem. Microbiology Australia 40:5–7

Chichlowski M, Guillaume De Lartigue J, Raybould HE, Mills DA (2012) Bifidobacteria isolated from infants and cultured on human milk oligosaccharides affect intestinal epithelial function. J Pediatr Gastroenterol Nutr 55(3):321–327

Chifiriuc MC, Gheorghe I, Czobor I, Florea DA, Mateescu L, Caplan ME, Caplan DM, Lazar V (2017) Advances in molecular biology-based assays for the rapid detection of food microbial contaminants. In: Food preservation. Academic Press, pp 645–669

Choudhary R, Singh KS, Bisht S, Kumar S, Mohanty AK, Grover S, Kaushik JK (2023) Host-microbe interaction and pathogen exclusion mediated by an aggregation-prone surface layer protein of Lactobacillus helveticus. Int J Biol Macromol 125146. https://doi.org/10.1016/j.ijbiomac.2023.125146. Epub ahead of print. PMID: 37271267

Cocolin L, Alessandria V, Dolci P, Gorra R, Rantsiou K (2013) Culture independent methods to assess the diversity and dynamics of microbiota during food fermentation. Int J Food Microbiol 167(1):29–43

Colomer-Lluch M, Jofre J, Muniesa M (2011) Antibiotic resistance genes in the bacteriophage DNA fraction of environmental samples. PLoS One 6(3):e17549. https://doi.org/10.1371/journal.pone.0017549

Critchley IA, Karlowsky JA (2004) Optimal use of antibiotic resistance surveillance systems. Clin Microbiol Infect 10(6):502–511. https://doi.org/10.1111/j.1469-0691.2004.00911.x

Doyle RM, Burgess C, Williams R, Gorton R, Booth H, Brown J, Bryant JM, Chan J et al (2018) Direct whole-genome sequencing of sputum accurately identifies drug-resistant mycobacterium tuberculosis faster than MGIT culture sequencing. J Clin Microbiol 56:e00666–e00618

Fisman DN, Laupland KB (2010) The 'One Health' paradigm: time for infectious diseases clinicians to take note? Canadian J Infect Dis Med Microbiol 21(3):111–114

Fittipaldi N, Segura M, Grenier D, Gottschalk M (2012) Virulence factors involved in the pathogenesis of the infection caused by the swine pathogen and zoonotic agent *Streptococcus suis*. Future Microbiol 7(2):259–279

FitzGerald RJ (1998) Potential uses of caseinophosphopeptides. s 8:451–457

FitzGerald RJ, Murray BA, Walsh DJ (2004) Hypotensive peptides from milk proteins. J Nutr 134: 980S–988S

Forghani F, Langaee T, Eskandari M, Seo KH, Chung MJ, Oh DH (2015) Rapid detection of viable Bacillus cereus emetic and enterotoxic strains in food by coupling propidium monoazide and multiplex PCR (PMA-mPCR). Food Control 55:151–157

Franco-Duarte R, Cernakova L, Kadam S, Kaushik KS, Salehi B, Bevilacqua A, Corbo MR, Antolak H, Dybka-Stępien K, Leszczewicz M, Relison Tintino S (2019) Advances in chemical and biological methods to identify microorganisms—from past to present. Microorganisms 7(5):130

Fuentes DN, Calderón PF, Acuña LG, Rodas PI, Paredes-Sabja D, Fuentes JA, Gil F, Calderón IL (2015) Motility modulation by the small non-coding RNA SroC in salmonella typhimurium. FEMS Microbiol Lett 362:135

Fugel HJ, Nuijten M, Postma M et al (2016) Economic evaluation in stratified medicine: methodological issues and challenges. Front Pharmacol 7:113

Fusco V, Quero GM (2014) Culture-dependent and culture-independent nucleic-acid-based methods used in the microbial safety assessment of milk and dairy products. Compr Rev Food Sci Food Saf 13(4):493–537

Gilbert JA, Dupont CL (2011) Microbial metagenomics: beyond the genome. Annu Rev Mar Sci 3: 347–371

Gobert G, Cotillard A, Fourmestraux C, Pruvost L, Miguet J, Boyer M (2018) Droplet digital PCR improves absolute quantification of viable lactic acid bacteria in faecal samples. J Microbiol Methods 148:64–73

Gong H, Vu GP, Bai Y, Chan E, Wu R, Yang E, Liu F, Lu S (2011) A salmonella small non-coding RNA facilitates bacterial invasion and intracellular replication by modulating the expression of virulence factors. PLoS Pathog 7:1002120

González Plaza JJ (2020) Small RNAs as fundamental players in the transference of information during bacterial infectious diseases. Front Mol Biosci 7:101

Greene AC (2018) CRISPR-based antibacterials: transforming bacterial defense into offense. Trends Biotechnol 36(2):127–130

Harris KA, Breaker RR (2018) Large noncoding RNAs in bacteria. Microbiol Spectr 6(4):515–526

He S, Han J, Lichtfouse E (2021) Backward transmission of COVID-19 from humans to animals may propagate reinfections and induce vaccine failure. Environ Chem Lett 19:763. https://doi.org/10.1007/s10311-020-01140-4

Herbel SR, Lauzat B, von Nickisch-Rosenegk M, Kuhn M, Murugaiyan J, Wieler LH, Guenther S (2013) Species-specific quantification of probiotic lactobacilli in yoghurt by quantitative real-time PCR. J Appl Microbiol 115(6):1402–1410

ISO (2005a) Microbiology of food and animal feeding stuffs—polymerase chain reaction (PCR) for the detection of foodborne pathogens—performance testing for thermal cyclers. International Standardization Organization, Geneva
ISO (2005b) Microbiology of food and animal feeding stuffs. General requirements and definitions. International Standardization Organization, Geneva, Polymerase chain reaction (PCR) for the detection of foodborne pathogens
ISO (2006a) Microbiology of food and animal feeding stuffs—polymerase chain reaction (PCR) for the detection of foodborne pathogens—requirements for amplification and detection for qualitative methods. International Standardization Organization, Geneva
ISO (2006b) Microbiology of food and animal feeding stuffs—polymerase chain reaction (PCR) for the detection of foodborne pathogens—requirements for sample preparation for qualitative detection. International Standardization Organization, Geneva
ISO (2011a) Microbiology of food and animal feeding stuffs—polymerase chain reaction (PCR) for the detection and quantification of foodborne pathogens—performance characteristics. International Standardization Organization, Geneva
ISO (2011b) Microbiology of food and animal feeding stuffs—real-time polymerase chain reaction (PCR) for the detection of foodborne pathogens—general requirements and definitions. International Standardization Organization, Geneva
Jani K, Dhotre D, Bandal J, Shouche Y, Suryavanshi M, Rale V, Sharma A (2018) World's largest mass bathing event influences the bacterial communities of Godavari, a holy river of India. Microb Ecol 76(3):706–718. https://doi.org/10.1007/s00248-018-1169-1
Jany JL, Barbier G (2008) Culture-independent methods for identifying microbial communities in cheese. Food Microbiol 25(7):839–848
Kuleshov V, Snyder MP, Batzoglou S (2016) Genome assembly from synthetic long read clouds. Bioinformatics 32:216–224
Kumar H et al (2015) Secretor status is strongly associated with microbial alterations observed during pregnancy. PLoS One 10:1–10
Kumar K, Cowley M, Davis R (2019) Next-generation sequencing and emerging technologies. Semin Thromb Hemost 45(7):661–673. https://doi.org/10.1055/s-0039-1688446
Kumbhare SV, Kumar H, Chowdhury SP et al (2017) A cross-sectional comparative study of gut bacterial community of Indian and Finnish children. *Sci Rep* **7**:10555. https://doi.org/10.1038/s41598-017-11215-y
Kuntz S, Rudloff S, Kunz C (2008) Oligosaccharides from human milk influence growth-related characteristics of intestinally transformed and non-transformed intestinal cells. Br J Nutr 99:62–471
Landt SG, Lesley JA, Britos L, Shapiro L (2010) CrfA, a small noncoding RNA regulator of adaptation to carbon starvation in Caulobacter crescentus. J Bacteriol 192:4763–4775
Lee YH, Kim S, Helmann JD, Kim BH, Park YK (2013) RaoN, a small RNA encoded within salmonella pathogenicity Island-11, confers resistance to macrophage-induced stress. Microbiology 159:1366
Liepke C, Adermann K, Raida M, Mägert HJ, Forssmann WG, Zucht HD (2002) Human milk provides peptides highly stimulating the growth of bifidobacteria.*Eur*. J Biochem 269:712–718
López-Posadas R, Requena P, González R, Suárez MD, Zarzuelo A, de Medina FS et al (2010) Bovine glycomacropeptide has intestinal anti-inflammatory effects in rats with dextran sulfate-induced colitis. *J Nutr* **140**:2014–2019
Luana FA, Cauã AW, Gabriel LL, Guilherme MVS, Tiago CB, María EG (2018) Metagenomic approaches for understanding new concepts in microbial science. International Journal of Genomics 2018:1–15. https://doi.org/10.1155/2018/2312987
Majdalani N, Chen S, Murrow J, St John K, Gottesman S (2001) Regulation of RpoS by a novel small RNA: the characterization of RprA. Mol Microbiol 39:1382–1394
Martínez JL, Coque TM, Baquero F (2015) Prioritizing risks of antibiotic resistance genes in all metagenomes. Nat Rev Microbiol 13(6):396–396. https://doi.org/10.1038/nrmicro3399-c2

Massé E, Gottesman S (2002) A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. Proc Natl Acad Sci 99:4620–4625

Matijasic BB, Obermajer T, Rogelj I (2010) Quantification of *lactobacillus gasseri, enterococcus faecium* and *Bifidobacterium infantis* in a probiotic OTC drug by real-time PCR. Food Control 21(4):419–425

Mattick JS (2001) Non-coding RNAs: the architects of eukaryotic complexity. EMBO Rep 2:986–991

McCartney AL (2002) Application of molecular biological methods for studying probiotics and the gut flora. Br J Nutr 88:29–37

Meng XC, Pang R, Wang C, Wang LQ (2010) Rapid and direct quantitative detection of viable bifidobacteria in probiotic yoghurt by combination of ethidium monoazide and real-time PCR using a molecular beacon approach. J Dairy Res 77(4):498–504

Metzker ML (2010) Sequencing technologies the next generation. Nat Rev Genet 11(1):31–46

Michelle S, Satola S, Read W, Timothy D (2018) Genome-based prediction of bacterial antibiotic resistance. J Clin Microbiol 57(3):e01405–e01418. https://doi.org/10.1128/JCM.01405-18

Mortari A, Lorenzelli L (2014) Recent sensing technologies for pathogen detection in milk: a review. Biosens Bioelectron 60:8–21

Muniandy S, The SJ, Thong KL, Thiha A, Dinshaw IJ, Lai CW, Ibrahim F, Leo BF (2019) Carbon nanomaterial-based electrochemical biosensors for foodborne bacterial detection. Crit Rev Anal Chem 49(6):510–533

Naarding MA, Ludwig IS, Groot F, Berkhout B, Geijtenbeek TB, Pollakis G, Paxton WA (2005) Lewis X component in human milk binds DC-SIGN and inhibits HIV-1 transfer to CD4+ T lymphocytes. J Clin Invest 115:3256–3264

Nakajima K, Tamura N, Kobayashi-Hattori K, Yoshida T, Hara-Kudo Y, Ikedo M, Sugita-Konishi-Y, Hattori M (2005) Prevention of intestinal infection by glycomacropeptide. Biosci Biotechnol Biochem 69:2294–2301

Nallari P (2020) Editorial on genomics innovations and advanced technologies. J Genomics Gene Study 3(3)

Naum M, Lampel KA (2016) DNA-based assays. In: Smithers GW (ed) Reference module in food

Nordmann P (2014) Carbapenemase-producing Enterobacteriaceae: overview of a major public health challenge. Med Mal Infect 44(2):51–56. https://doi.org/10.1016/j.medmal.2013.11.007

Ogier JC, Son O, Gruss A, Tailliez P, Delacroix-Buchet A (2002) Identification of the bacterial microflora in dairy products by temporal temperature gradient gel electrophoresis. Appl Environ Microbiol 68(8):3691–3701

One Health (2008) A new professional imperative. American Veterinary Medical Association, p 9. Retrieved 2017-08-20

O'Rourke A, Beyhan S, Choi Y, Morales P, Chan AP, Espinoza JL, Dupont CL, Meyer KJ, Spoering A, Lewis K, Nierman WC (2020) Mechanism-of-action classification of antibiotics by global transcriptome profiling. Antimicrob Agents Chemother 6:1207–1219

O'Sullivan DJ, Giblin L, McSweeney PLH, Sheehan JJ, Cotter PD (2013) Nucleic acid-based approaches to investigate microbial-related cheese quality defects. Front Microbiol 4:1

Pace NR, Stahl DA, Lane DJ, Olsen GJ (1986) The analysis of natural microbial populations by ribosomal RNA sequences. In: Advances in Microbial Ecology, pp 1–55

Pellegrini A, Thomas U, Bramaz N, Hunziker P, Fellenberg V (1999) Isolation and identification of three bactericidal domains in the bovine α-lactalbumin molecule. Biochim Biophys Acta 1426: 439–448

Perumal V, Hashim U (2014) Advances in biosensors: principle, architecture and applications. J Appl Biomed 12(1):1–15

Phillips KA, Deverka PA, Marshall DA, Wordsworth S, Regier DA, Christensen KD, Buchanan J (2018a) Methodological issues in assessing the economic value of next-generation sequencing tests: many challenges and not enough solutions. Value Health 21(9):1033–1042. https://doi.org/10.1016/j.jval.2018.06.017

Phillips KA, Deverka PA, Hooker GW et al (2018b) Genetic test availability and spending: where are we now? Where are we going? Health Aff (Millwood) 37:710–716

Pineda M, Lear A, Collins JP, Kiani S (2019) Safe CRISPR: challenges and possible solutions. Trends Biotechnol 37:389–401

Porcellato D, Narvhus J, Skeie SB (2016) Detection and quantification of *Bacillus cereus* group in milk by droplet digital PCR. J Microbiol Methods 127:1–6

Qian L, Song H, Cai W (2016) Determination of bifidobacterium and lactobacillus in breast milk of healthy women by digital PCR. Benef Microbes 7(4):559–569

Rajapaksha P, Elbourne A, Gangadoo S, Brown R, Cozzolino D, Chapman J (2019) A review of methods for the detection of pathogenic microorganisms. Analyst 144(2):396–411

Rajendhran J, Gunasekaran. (2011) Microbial phylogeny and diversity: small subunit ribosomal RNA sequence analysis and beyond. Microbiol Res 166(2):99–110

Randazzo CL, Caggia C, Neviani E (2009) Application of molecular approaches to study lactic acid bacteria in artisanal cheeses. J Microbiol Methods 78(1):1–9

Razzaq A, Masood A (2018) CRISPR/Cas9 system: a breakthrough in genome editing. Mol Biol 7: 2

Repoila F, Darfeuille F (2009) Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. Biol Cell 101:117–131

Rider TH, Petrovick MS, Nargi FE, Harper JD, Schwoebel ED, Mathews RH, Blanchard DJ, Bortolin LT, Young AM, Chen J, Hollis MA (2003) AB cell-based sensor for rapid identification of pathogens. Science 301:213–215

Rodgers K, McLellan I, Peshkur T et al (2019) Can the legacy of industrial pollution influence antimicrobial resistance in estuarine sediments? Environ Chem Lett 17:595–607. https://doi.org/10.1007/s10311-018-0791-y

Rubab M, Shahbaz HM, Olaimat AN, Oh DH (2018) Biosensors for rapid and sensitive detection of Staphylococcus aureus in food. Biosens Bioelectron 105:49–57

Sabat AJ, Budimir A, Nashev D, Sá-Leão R, van Dijl J, Laurent F, Grundmann H, Friedrich AW (2013) Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill 18(4):20380

Sai-Anand G, Sivanesan A, Benzigar MR, Singh G, Gopalan AI, Baskar AV, Ilbeygi H, Ramadass K, Kambala V, Vinu A (2019) Recent progress on the sensing of pathogenic bacteria using advanced nanostructures. Bull Chem Soc Jpn 92(1):216–244

Salina EG, Grigorov A, Majorov K, Logunova N, Ignatov D, Skvortsova Y, Bychenko O, Apt A, Kaprelyants A, Azhikina T (2018) DrrS, a small non-coding mycobacterium tuberculosis RNA, regulates the whole genome expression shifts consistent with adaptations for survival within host macrophages. bioRxiv:474304

Schmieger H, Schicklmaier P (1999) Transduction of multiple drug resistance of salmonella enterica serovar typhimurium DT104. FEMS Microbiol Lett 170(1):251–256. https://doi.org/10.1111/j.1574-6968.1999.tb13381.x

Schoenfelder SM, Lange C, Prakash SA, Marincola G, Lerch MF, Wencker FD, Förstner KU, Sharma CM, Ziebuhr W (2019) The small non-coding RNA RsaE influences extracellular matrix composition in *Staphylococcus epidermidis* biofilm communities. PLoS Pathog 15: 1007618

Schwarze K, Buchanan J, Taylor JC et al (2018) Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. Genet Med 20(10):1122–1130. https://doi.org/10.1016/b978-0-08-100596-5.00597-7

Sen R, Nayak L, De RK (2016) A review on host–pathogen interactions: classification and prediction. Eur J Clin Microbiol Infect Dis 35:1581–1599. https://doi.org/10.1007/s10096-016-2716-7

Sengupta R, Altermann E, Anderson RC, McNabb WC, Moughan PJ, Roy NC (2013) The role of cell surface architecture of lactobacilli in host-microbe interactions in the gastrointestinal tract. Mediat Inflamm 2013:237921. https://doi.org/10.1155/2013/237921

Sharma A, Lee S, Park YS (2020) Molecular typing tools for identifying and characterizing lactic acid bacteria: a review. Food Sci Biotechnol 29(10):1301–1318

Shen ZQ, Wang JF, Qiu ZG, Jin M, Wang XW, Chen ZL, Li JW, Cao FH (2011) QCM immunosensor detection of Escherichia coli O157: H7 based on beacon immunomagnetic nanoparticles and catalytic growth of colloidal gold. Biosens Bioelectron 26(7):3376–3381

Shepelkova G, Evstifeev V, Averbakch M Jr, Sivokozov I, Ergeshov A, Azhikina T, Yeremeev V (2021) Small noncoding RNAs MTS0997 and MTS1338 affect the adaptation and virulence of mycobacterium tuberculosis. Microbiol Res 12:186–195

Singh K et al (2017) Expression of recombinant truncated domains of mucus-binding (Mub) protein of lactobacillus plantarum in soluble and biologically active form. Protein Expr Purif 135:54–60

Singh KS, Kumar S, Mohanty AK et al (2018) Mechanistic insights into the host-microbe interaction and pathogen exclusion mediated by the Mucus-binding protein of *Lactobacillus plantarum*. Sci Rep 8:14198. https://doi.org/10.1038/s41598-018-32417-y

Singh KS, Anand S, Dholpuria S et al (2021) Antimicrobial resistance dynamics and the one-health strategy: a review. Environ Chem Lett 19:2995–3007. https://doi.org/10.1007/s10311-021-01238-3

Smith CJ, Osborn AM (2009) Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. FEMS Microbiol Ecol 67:6–20

Sofos JN (2008) Challenges to meat safety in the 21st century. Meat Sci 78:3–13

Sohier D, Pavan S, Riou A, Combrisson J, Postollec F (2014) Evolution of microbiological analytical methods for dairy industry needs. Front Microbiol 7(5):16

Sudhakaran VA, Anand S (2020) Approaches for detection of dairy microorganisms: an update. In: Dairy processing: advanced research to applications, pp 217–241

Tasan I, Zhao H (2017) Targeting specificity of the CRISPR/Cas9 system, vol 6, p 1609

Tringe SG, Hugenholtz P (2008) A renaissance for the pioneering 16S rRNA gene. Curr Opin Microbiol 11:442–446

Urashima T, Taufik E (2011) Oligosaccharides in milk: their benefits and future utilization. Media Peternakan 33(3):189–197

Wachter S, Hicks LD, Raghavan R, Minnick MF (2020) Novel small RNAs expressed by *Bartonella bacilliformis* under multiple conditions reveal potential mechanisms for persistence in the sand fly vector and human host. PLoS Negl Trop Dis 14:0008671

Wada Y, Lönnerdal B (2014) Bioactive peptides derived from human milk proteins—mechanisms of action. J Nutr Biochem 25(5):503–514. https://doi.org/10.1016/j.jnutbio.2013.10.012

Wallace JG, Zhou Z, Breaker RR (2012) OLE RNA protects extremophilic bacteria from alcohol toxicity. Nucleic Acids Res 40:6898–6907

Wang Z, Xu H, Wei Z, Jia Y, Wu Y, Qi X, Li Y, Gao X (2020) The role of non-coding RNA on macrophage modification in tuberculosis infection. Microb Pathog 149:104592

Wassarman KM, Storz G (2000) 6S RNA regulates *E. coli* RNA polymerase activity. Cell 101:613–623

Wetterstrand, K.A. 2018. DNA sequencing costs: data from the NHGRI genome sequencing program (GSP). https://www.genome.gov/sequencingcostsdata. Accessed 12 Dec 2018

Wordsworth S, Brett D, Katherine P, James B, Deborah M, McCabe A, Regier DA (2018) Using "Big Data" in the cost-effectiveness analysis of next-generation sequencing technologies: challenges and potential solutions. Value Health 21(9):1048–1053. https://doi.org/10.1016/j.jval.2018.06.016

Zampetaki A, Albrecht A, Steinhofel K (2018) Long non-coding RNA structure and function: is there a link? Front Physiol 9:1201

Zhang S, Liu S, Wu N, Zhang W, Zhang Y (2018) Identification of small regulatory RNAs involved in persister formation. bioRxiv:310631

Chapter 9
Methods for Measurement of Microbial Diversity

Vikas Sangwan, Namita Ashish Singh, Meena Sindhu, Arundhati Wandhare Ganesh, Anamika Singh, and Manvesh Kumar Sihag

Abstract The microbial population represents the richest biodiversity resources on the planet and the accurate measurement of microbial diversity is of utmost importance as they play a very pivotal role in the maintenance of life on earth. Measurement of diversity is very challenging specifically in the case of microorganisms. Due to their extremely small size and clonal nature, microorganisms pose very distinct measurement challenges. The rate at which they grow, evolve, and respond to their environment further increases the complications in diversity studies. Our ability to understand the microbial ecosystem is hampered by complicated cultivation and methodological limitations. Earlier, the study of microbiology is limited to conventional culture-based methods and microscopic observations. Lately, we have seen phenomenal progress in the development of technologies for the measurement of microbial diversity. Recent technological advancements along with powerful bioinformatics platforms have accelerated microbial diversity research globally. This chapter introduces various types of microbial diversities, their characterization, and the different conventional and latest technologies used in their measurement.

Keywords Microbial diversity · Next-generation sequencing · Flow cytometry

V. Sangwan (✉) · A. W. Ganesh · A. Singh
Department of Dairy Microbiology, College of Dairy Science and Technology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

N. A. Singh
Department of Microbiology, Mohanlal Sukhadia University, Udaipur, Rajasthan, India

M. Sindhu
Department of Microbiology, College of Basic Sciences and Humanities, Chaudhary Charan Singh Haryana Agricultural University, Hisar, Haryana, India

M. K. Sihag
Department of Dairy Chemistry, College of Dairy Science and Technology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_9

9.1 Introduction

Biological diversity, most commonly known as biodiversity in modern science, is explained as the range of variation among some set of attributes. Biological diversity thus refers to the sum total of all the biotic variation within the living world from the genetic to ecosystem level (Purvis and Hector 2000). The term biodiversity was earlier used in the context of the number of species but by using various conventional methods of identification, it was quite difficult to separate different species, various other elaborated definitions of biodiversity were described including the variability and variety of the living organisms. These simplified and elaborated definitions of biodiversity are well debated over a period of time taking into consideration the various scientific aspects, national and international agreements, and different conservation initiatives. Even after long going global discussions, we do not have a universally agreed definition of biodiversity. Generally, while defining the term biodiversity, the concept of genes, ecosystem, and species is usually considered. In the simplest form, biodiversity can be explained as a number of species in a given area. But this explanation was rejected by different researchers with a notion that biodiversity is much more than the number of different species (Swingland 2013). The variety within and among living species, assemblages of living organisms, biotic communities, and biotic processes—whether naturally occurring or altered by humans are all included in the broad definition of biodiversity (DeLong 1996). Biodiversity, in its broadest definition, includes all heritably based variation at all levels of organizational structure, from genes through populations, and species to communities and ecosystems.

The key objective of ecology is to explain and predict the biodiversity patterns across evolutionary distant taxa. Generally, the concept of biodiversity is explained with reference to the large or macroorganisms, but in terms of their abundance, diversity, and ubiquitous nature, microorganisms are far ahead of every other organism on this earth. Initially, for a very long time, the focus of evolutionary studies was focused mainly on the eukaryotic plants and animals with very little attention paid to the microorganisms. However, global efforts to study microbial diversity across different ecosystems have created an opportunity to understand biodiversity using a scale of data that surpasses the datasets of macroorganisms (Shoemaker et al. 2017). Microbes are tiny living creatures that inhabit practically all of nature's habitats, like soil, dunes, oceans, and rocks. Millions of microbes reside in and on our body including skin, oral cavity, and gastrointestinal tract. Bacteria, archaea, eukaryotes, and viruses comprise the four major categories of microorganisms. According to a recent estimate, there are over one trillion different species of microorganisms on the planet, of which 99.999 percent have not yet been identified. These bacteria create communities with astounding diversity in their hosts or habitats (Maiden et al. 2013). Microorganisms were the first cellular inhabitant of the earth and were active here billions of years before the appearance of macroorganisms. Their metabolic activities played a pivotal role in creating an environment that is conducive to the growth and evolution of multicellular

organisms. Initially, in the absence of the latest techniques for the study of biodiversity, the establishment of the evolutionary relationship between organisms was very challenging and organisms were divided into prokaryotes and eukaryotes, which were distinguished by the presence or absence of defined nuclear membrane. After the development of molecular methods, the information provided by studying the sequence of nucleic acids revealed three very distinct domains; bacteria, archaea, and eukarya. Morphologically, bacteria and archaea are quite alike but archaea are more similar to eukaryotes in many genetic and cellular characteristics. The quantification of microbial biodiversity is very different as well as difficult in comparison to the macroorganisms because of their very small size and clonal nature. Only a small fraction of microbial diversity has been described to date. Although it is very difficult or nearly infeasible to completely quantify the diversity in microorganisms in a particular habitat, a number of studies estimate microbial diversity in different habitat types (Matulich et al. 2013). This chapter aims to impart information about microbial diversity including types of microbial diversity, its measurement, and most commonly used techniques to measure microbial biodiversity.

9.2 Microbial Diversity

Microorganisms are extremely diverse with respect to their structure, function, metabolism, shape, size, and arrangement of cells. Based on these parameters, microbial diversity can be broadly classified based on their morphology, structure, metabolism, ecology, and behavior.

9.2.1 Morphological Diversity

Being very small and comparatively simple than higher organisms, microbial cells might be expected to have a uniform shape and size. But the microbial world offers considerable variety in terms of morphology. The structural framework of the cell is the basic essence of any living organism and includes appearance, form, and the visually recognizable characters. All these features are collectively known as morphology. Microorganisms are often described in terms of their shapes. Along with shape, different groups of microorganisms also differ in the arrangement of their cells. Shapes and arrangement of cells affect important biological functions, including motility, stress resistance, nutrient uptake, and interaction with other organisms, which also add up to the diversity. Bacteria are extremely diverse in morphology. The cell wall of bacteria allows them to maintain a definite shape. Two principal aspects of bacterial cells are arrangement and shape. As far as the bacterial cell arrangement is concerned, it may be single, paired (*Streptococcus pneumoniae*), cluster (*Staphylococcus aureus*) or chains (*Streptococcus thermophilus*), palisade (*Corynebacterium diphtheriae*), and tetrad (*Pediococcus*). In shape, they may

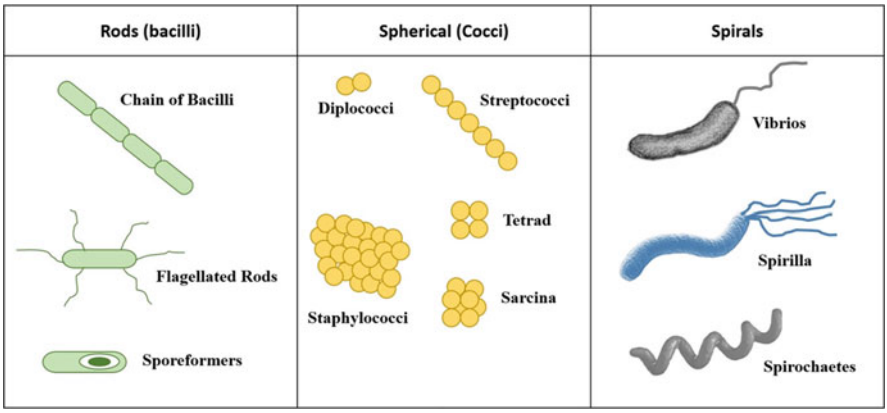

Fig. 9.1 Morphological diversity in bacteria

principally be rods (bacilli), spheres (cocci), and spirals (spirillum). Most common bacterial shapes are represented in Fig. 9.1. Along with these common shapes, some cells also have other shapes including filaments (*Candida tussavagella*), star-shaped (*Stella*), rectangular (*Haloarcula vallismortis*), and many others (van Teeseling et al. 2017). Some bacteria are even considered pleomorphic (*Mycoplasma pneumoniae*), i.e., they do not have any characteristic shape, unlike all others. In pure cultures, they can be observed to have different shapes. The morphological diversity of eukaryotic cells is one of their distinguishing features. This is especially true for eukaryotic microbes, which can be found in a variety of environments and have evolved morphological adaptations accordingly. As bacterial cells, fungal and viral cells are highly diverse with respect to their shapes, size, and structures. The kingdom Fungi consist of highly diverse heterotrophic eukaryotic microorganisms with a range of cell organizations including simple unicellular cells to highly complex filamentous structures. The members of this kingdom are characterized by cell wall made of chitin, and lack of phagotrophic capabilities (Naranjo-Ortiz and Gabaldón 2019). Fungi vary widely in size and shape; they may be unicellular and microscopic or may have multicellular forms which can easily be seen with the naked eye. The size of the individual cells ranges from 1 μm to 30 μm. Microscopic fungi exist as either molds or yeasts or both. Fungi, also differ in internal structure, some have septate hyphae (long, branching filamentous structure of a fungus which acts as the main mode of vegetative growth and is collectively known as mycelium), while others have nonseptate hyphae. Different fungi have different modes of reproduction and produce different types of spores like basidiospores (*Agaricus bisporus*) and ascospores (*Penicillium chrysogenum*). Similar kind of morphological diversity is also represented by viruses. Viruses vary in their shapes (rod, helical, spherical, icosahedral, pentagonal, and hexagonal), structure, and genetic materials (DNA and RNA).

9.2.2 Structural Diversity

All the different type of microbes differ from the other in various ways, with one difference being structure. Despite having some similar traits, cells can differ greatly. Prokaryotic cells (lack membrane-bound nucleus) and eukaryotic cells (contain well-organized and membrane-bound nuclei) are the two primary types of cells. There is tremendous diversity in the structure, function, and metabolic activity of each type of cell. Although viruses are biological entities with some similarities to cells, they are not cells. In addition to having a vast range of shapes and reproduction techniques, viruses also need a host cell to reproduce since, unlike cells, they do not metabolize. The cytoskeleton, cell membrane, and/or cell wall are only a few examples of the numerous structures that may support the wide variety of morphologies found in eukaryotic cells. Prokaryotic cells are distinct from eukaryotic cells in that they have a nucleoid, rather than a membrane-bound nucleus, as the location of their genetic material. Prokaryotic cells typically do not have membrane-bound organelles. The majority of prokaryotic cells contain a cell wall that aids in maintaining cellular shape and shields the organism from osmotic pressure changes. The cell envelopes of many bacteria are typically gram-positive or gram-negative ones. However, even within these two main kinds, lots of variations are there. The outer membrane of many gram-positive bacteria is comprised of mycolic acids rather than glycerol-phosphate esters. The lipopolysaccharide layer is absent in many gram-negative bacteria. The S-layer, an ordered protein coat, is present on numerous bacteria and archaea (Brown 2014). Peptidoglycan makes up the cell walls of bacteria; however, this type of substance has an unexpectedly wide range of chemical changes. Although some archaeal cell walls contain a substance similar to peptidoglycan, called pseudomurein, archaea do not have peptidoglycan cell walls.

9.2.3 Metabolic Diversity

Microorganisms acquire the energy and nutrients it needs to survive and reproduce through metabolism. Microbes employ a wide variety of metabolic techniques, and metabolic differences across species are frequently used to distinguish between them. A microbe's ecological niche is mostly determined by its particular metabolic characteristics, which frequently enable the organism to be beneficial in industrial operations or responsible for biogeochemical cycles. Metabolic diversity in macroscopic eucaryotes is very limited in comparison to bacteria and archaea. Macroscopic eukaryotes can be either chemoheterotrophic (e.g., animals) or photoautotrophic (e.g., plants), while bacteria and archaea have a very wide range of sources to extract energy and carbon. Metabolic types of bacteria are generally defined by the source of energy, carbon, and electron. For energy, only two sources are available including light and some chemical molecules. Organisms using light as an energy source are known as phototrophs while those using chemical compounds

are called chemotrophs. On a similar pattern, microorganisms have only two types of electron sources. Those using inorganic substances as electron sources are termed lithotrophs and organisms using organic compounds are known as organotrophs. On the basis of carbon sources, organisms can be divided into heterotrophs (carbon source—reduced organic molecules) and autotrophs (carbon source—carbon dioxide). By combining these metabolic types, both bacteria and archaea can be divided into five broad nutritional groups; photolithoautotrophs (purple and green sulfur bacteria), photoorganoheterotrophs (purple nonsulfur bacteria), chemolithoautotrophs (methanogens, nitrifying bacteria), chemolithoheterotrophs (Sulfur-oxidizing bacteria), and chemoorganoheterotrophs (most nonphotosynthetic microbes).

9.2.4 Ecological Diversity

Microbes are thought to live everywhere on the earth, even in an environment that is considered inhospitable also termed extreme environments. As all bacteria, archaea, and most eukarya lineages are microorganisms, the microbial world contains the majority of the earth's phylogenetic diversity. Microbes live in all kinds of environments, and their presence always has an impact on the environment in which they grow. Their adaptability allows them to thrive in extreme environments. Microorganisms are inhabitants of a wide range of habitats, from freshwater sources to oceans, ice, hot water springs, salt lakes, and places with a highly acidic environment. Microbes can also grow at a wide range of temperatures with highly variable optimum temperatures from 0 °C to 113 °C. Microorganisms can also be classified based on their ecological niches divided on the basis of solute concentration (halophiles, osmotolerant, xerophiles), pH (acidophiles, neutrophiles, and alkaliphiles), temperature (psychrophiles, psychrotolerant, mesophiles, thermophiles, hyperthermophiles), oxygen concentration (obligate aerobe, facultative anaerobe, aerotolerant anaerobe, obligate anaerobe, microaerophiles), pressure (piezophile), and radiation-resistant. Also in these environments, microorganisms behave differently and exhibit different types of neutral, positive, and negative interactions with each other. However, not every microbe can survive in all the environmental conditions mentioned above. Microorganisms are very specific for the conditions they live in. It is generally acknowledged that the impact of microorganisms on their environment can be advantageous, despite the fact that the vast majority of the diversity of microbes is yet unknown. The metabolic activities of these organisms are mostly responsible for the positive effects of microorganisms, such as their interactions with plants and animals, their usage in food production, and biotechnological processes.

9.2.5 Behavioral Diversity

As much as it seems odd to believe that these small microorganisms have social behavior, they really do have behavior. The adaptive behavior shown by microorganisms in response to a specific stimulus is also known as microbial intelligence. The concept of behavioral intelligence includes complicated adaptive behavior exhibited by a single cell as well as altruistic or cooperative conduct in populations of similar or dissimilar cells, both of which are influenced by chemical signaling that alters cellular behavior or physiology and affects colony structures. Algae and protozoa are examples of complex cells that exhibit exceptional behavioral skills. Amoeba's ability to construct shells demonstrates sophisticated manipulation and discrimination abilities, which are typically assumed to only be present in multicellular organisms. Even bacterial populations can behave differently. These actions take place in populations of just one species or a mix of species. Myxobacteria colonies or swarms, quorum sensing, and biofilms are a few examples (Chimileski and Kolter 2017). The responses of microorganisms toward their surrounding environment are not predictable and are generally influenced by the physiological state of the cells. Motility and taxis are types of behavior that come in a variety of forms. All motile microorganisms respond to a variety of chemical stimuli.

Different types of microorganisms respond variably to environmental stimuli, some respond to changes in light intensity, while some respond to magnetic fields, solute, and nutrient concentration. Motile microbes utilize the thread-like appendages known as flagella for swimming, moving, or gliding. The responses toward the environmental cues are basically due to the change in the intensity of a stimulus. Microorganisms can sense the variation in the intensity of stimulus over a period of time as they have a membrane potential-based working memory with a time constant of 0.5–1 s (Fenchel 2002; Yang et al. 2020).

The tactic behavior of microorganisms ranges from the use of gas vacuoles for regulation of their position in the water by phototactic bacteria (*Cholorobium* spp.) to chemotactic bacteria (*Rhizobium* spp.), which senses various chemical signals from plant roots and swim by using flagella. Numerous bacteria have the capacity to alter their swimming pattern in response to alterations in the environment, moving toward conditions that are more suitable for growth. It has become clear that the same network of cytoplasmic proteins coordinates the response to a variety of stimuli, including light, temperature, and nutrient concentrations, and that the essential elements of this network are constant across bacterial species. Magnetotactic bacteria have an internal magnetic compass that allows them to navigate the Earth's magnetic field.

Every organism goes through many stages of development; at the very least, they can move from an actively developing log phase to a resting or slowly growing stationary phase. Other developmental cycles include sporulation, the production of swarmer cells, cysts, and terminal differentiation into germ and somatic cells. Examples of these processes include heterocysts in cyanobacterial filaments,

"slugs" in myxobacteria, and the incredibly complex life cycles of *Streptomyces* species. The metabolic responses of microbes to their surroundings include the expression of the genes required to compete for the resources present at the time. Switching of metabolism from oxidative to fermentative after complete exhaustion of oxygen or utilization of less preferred sugar after complete utilization of primary sugar (galactose in place of glucose) are some examples of this behavior. Apart from showing these behavioral aspects, microbes also act communally and in order to communicate, organisms can send and receive chemical signals or make physical touch. As an illustration, *Myxococcus* swarming starts with a chemical signal that spreads throughout the community and draws the cells together with the help of multiple signaling and special polysaccharides (Islam et al. 2020). The aggregation and development of fruiting bodies are then driven by direct contact between cells. Additionally, certain microorganisms develop symbiotic relationships with other microbes or with larger organisms. Complex microbial communities form "mats" that process and recycle nutrients for the community as a whole. Individual cells of protists coordinate to produce complex structures or move as multicellular entities. The single-celled ciliate *Stentor roeselii* exhibits a form of "behavioral hierarchy" and expresses the ability to "alter its mind" when responding to an irritant that does not go away, suggesting a very speculative concept of cognition (Tang and Marshall 2018).

9.3 Characterization of Microbial Diversity

Due to their immense diversity and small size, the quantification of biodiversity in microorganisms is very challenging in comparison to macroorganisms. Although establishing a macroorganism species can be difficult, determining a microbial species is extremely difficult, in part due to their genetic makeup. Microorganisms are also too small to be classified by morphological traits, which are often done by taxonomists of plants and animals. Even with the help of highly advanced techniques, it is not feasible to quantify microbial diversity in even a gram of soil, milk, and marine water. The diversity studies of microorganisms focus on variations in DNA rather than their phenotypes in stark contrast to the diversity studies of large organisms. Two crucial criteria for determining the diversity and structure of the community are evenness and richness of species which are described as the population size of a particular species within a community and the number of species within a community respectively. Because a significant portion of the organisms found in nature seems to be resistant to cultivation, conventional culture-dependent approaches have a very limited ability to quantify these two properties. The fact that changing the environmental conditions during cultivation can change the community structure is another downside of these methods. As a result, a new community structure emerges that could not exactly correspond to the original structure (Liu et al. 1997). Today, a number of molecular methods offer an edge over the culture-dependent methods for the characterization of microbial diversity. The method that

combines PCR and rRNA-based phylogeny has proved successful in identifying uncultured organisms and exploring microbial habitats. This method includes isolation of community DNA, followed by gene amplification using PCR and the creation of a clone library for rDNA genes. The amount and variety of different clones in a community are used to determine its richness and evenness. While this method eliminates the need for cultivation, clone library development and screening are time-consuming. More recently, methods to avoid library creation have been proposed that rely on single-strand DNA conformation or DNA melting behavior. These techniques have been used to evaluate community structure and offer fair estimates of species diversity.

9.4 Measurement of Microbial Diversity

The approach of measuring the biodiversity in animals and plants by counting and identifying individual species is neither practical nor feasible in the case of microorganisms. Individual microorganisms can be counted by using microscopes but the morphological observation after microscopy does not actually reflect their diversity. Some characteristics of microorganisms that are used for classification remain conserved from generation to generation but relying on phenotypic information for classification sometimes leads to inaccuracy as a process like horizontal gene transfer is very common among microorganisms. Also, not all the methodologies used for the characterization of microorganisms end in species identification. Therefore, it is debatable whether the notion of species, which is used to categorize complex eukaryotes (plants and animals), can be extrapolated to microbes. While studying the biodiversity in the microbiological population, another important concern is to decide whether the target is taxonomic diversity or functional diversity. Because, in the case of ecological studies, the important actions of microbes like elemental transformation, organic residues decomposition, and biodegradation of environmental pollutants are of greater relevance than other less important structural information. Based on the observation that there are many more taxa than processes in most communities, it is hypothesized that there is a significant level of functional redundancy among microorganisms. In studies of ecological biodiversity, specific functions have thus far proved challenging to quantify and comprehend, particularly by nonmicrobiologists. However, the introduction of quick molecular assays has made it easier to measure the makeup of microbial communities, and taxonomic characterization is now more frequently utilized. It is important to compile functional and taxonomic diversity estimations for species represented by recently found ssrDNA sequences in order to better understand the ecological functions they perform. Currently, methods for the characterization of microbial communities are available in abundance. Based on the objectives of the study, the most important and crucial step is to identify which method is to be used for diversity studies.

9.5 Methods for Measurement of Microbial Diversity

As the number of available methods for the determination of microbial diversity is very high, not all methods can be explained in one place. The most commonly used methods are described here. The methods for measurement of microbial diversity described here are divided into three broad categories including cell count-based methods, process or function-based methods, and cellular constituent-based methods.

9.5.1 Cell Count-Based Methods

The quality of microbial diversity study is directly proportional to the method applied to examine and differentiate the different types of microorganisms present. Most of the diversity studies start with basic techniques which include counting and examining cells with the help of a microscope and traditional plate methods.

9.5.1.1 Plate Counts

Plate count is the most traditional method for the measurement of microbial diversity. In this method, viable and culturable microbes can be grown over solid nutrient media and enumerated. It is mainly used to study morphological characteristics of microbes such as colony shape, size, color, pigment production, etc. It is mainly used for the primary screening of different microbial species. Although it is inexpensive, there are certain limitations such as finding specific media, optimum temperature, and pH for the growth of microbes. Many bacteria and fungi produce inhibitory substances limiting the growth of each other and fast-growing species overcome the growth of slow-growing species. A smaller cell having a size less than 0.065 mm^3 cannot be cultivated using this method. Moreover, all the microbes are not culturable, only 1% of microbes can be grown using the agar plate method, so total measurement of microbial diversity cannot be done using this method (Nèble et al. 2007). Different stages of bacterial growth also affect their isolation in plate count as cells during the stationary phase are starved of nutrients and cannot be cultivated while the cells during the lag phase can grow (Winding et al. 2005; Oliver et al. 2005).

A significant portion of uncultured microorganisms—up to 80%, have the genetic makeup of the active bacterial population (Roszak and Colwell 1987; Bernard et al. 2007). This is the main reason for the underestimation of microbial biomass by the plate count method. Therefore, the isolation of microorganisms on agar or liquid medium is currently being used for the assessment of any specific activity and function (enzymatic activity), morphological identification, and physiological characteristics of microbial communities (Dunfield and Conrad 2000; Constant et al.

2008, 2010). Even though these traditional methods have some shortcomings, the colony characteristics were used for the determination of microbial growth and rates by utilizing the kinetic approach (Hashimoto and Hattori 1989).

The moment when a colony first appears on a microbiological growth media depends on (1) the duration of the lag phase; (2) the size of the microbial cell; and (3) microbial growth rates. Microbes that come from a nutritive media in an active physiological state generally have a shorter lag phase. Thus, in cases, where growth is not limited, identification of the fastest-growing microbial population on the growth media can be achieved by the colony-forming curves. The plate count estimates the total number of active microorganisms without defining the currently active microbial proportion; since cell division starts considerably earlier than the colony may be observed (Besset-Manzoni et al. 2018). In the analysis time required for colony formation, copiotrophic organisms (characterized as gram-positive and spore-forming organisms) were identified as the fastest-growing organisms (Kasahara and Hattori 1991). The cell sizes of potentially active bacteria are larger, but they typically contain 2.5 times fewer cells than slow-growing or dormant bacteria.

9.5.1.2 Microscopy

Microscopic study can be used for counting of cells and for observation of cell division. Complementary stains can be used for distinguishing living microbial cell from nonliving cells. Some dyes can cross cell membrane and have ability to bind nucleic acid and proteins such as SYBR Green I, acridine orange, 4,6-diamidino-2-phenylindole [DAPI], while other dyes which are not able to cross cell membranes are used to stain dead cell (Busse et al. 2009; Luna et al. 2004). Active cells are also estimated by subtracting dead cells from total microbial count using double staining with propinium but counting of dormant cell leads to overestimation which can be misnomer. Metabolically active cell can also be stained with fluorescent dyes, e.g., microbes having active respiration convert fluorescein diacetate into green compound and produce red colored formazan by reducing 5-cyano-2,3-ditolyl-tetrazolium chloride (CTC) or 2-(p-iodophenyl)-3-(p-nitrophenyl)-5-phenyl tetrazolium chloride (Maraha et al. 2004). Autoradiography can also be used for counting of living microbes, i.e., bacteria, fungi, and functional groups present on their surface (Stiehl-Braun et al. 2011) utilizing radioactive substrates (e.g., $_{3}H$- or $_{14}C$-glucose, $_{14}CH_4$). Problem with this technique is that fungal hyphae are considered dormant as no metabolic activity is shown by hollow hyphae (Heaton et al. 2012). In situ microbial activity may change during extraction from its sources. Enrichment of substrate also leads to increase in active number of microbes, while under stress conditions, the microbial population reduces which leads to change in microbial population (Maraha 2007).

9.5.2 Function-Based Method

Microbes make earth inhabitable through various biogeochemical cycles. Microbes play important role in all aspects of life, such as degradation and decomposition of dead and waste material, bioremediation, contribution in various nutrient cycles (carbon, nitrogen, phosphorus, sulfur), pathogenic agent, biocontrol, biofertilizer, regulation of climate, etc. (Ducklow 2008; Giller et al. 2009). All the ecological processes are influenced by microbes either directly or indirectly. Microbes are most diverse community of the nature and interact with all living communities such as plant, animals, and human beings (Koskella et al. 2017; McFall-Ngai 2015). As microbes are involved in various physiological activities, so based on their function on ecological niche, diversity of microbial cell can be quantified.

9.5.2.1 Enzyme Assay

Production of enzyme is the main characteristics of active microbial cell. Microorganisms such as bacteria, fungi, and actinomycetes produce various kinds of enzymes which include both extracellular and intracellular enzymes. Total enzymatic activity includes both extracellular as well as immobilized enzyme. Extracellular enzyme includes fluorescein diacetate hydrolysis, phosphomonoesterases, arylsulfatase, urease, and β-glucosidase while intracellular enzyme includes dehydrogenase. Extracellular enzyme can be associated with active cell, dead cell, and can form complex with humus and clay, while dehydrogenase is internal, so, accurate measurement for oxidative activity of microbes can be carried out by calculating dehydrogenase activity. These enzymes are essential for biogeochemical cycling and mineralization of carbon, hydrogen, nitrogen, and sulfur. Microbial diversity using enzyme can be studied by genomics and proteomics. Microbe produces various hydrolytic enzymes for degradation of the substrate.

The most prevalent enzymatic method for estimation of microbial diversity is the fluorescein diacetate (FDA) hydrolysis method. This method is used to determine the diversity of microbes such as bacteria, fungi, and actinomycetes among different environments such as freshwater, snow, and marine environment of microorganisms. Microbial enzymes such as protease, esterase, and lipase hydrolyze FDA into fluorescein, a colored product that is measured spectrophotometrically to estimate the active microbial population in the sample. But it is not accurate measure as hydrolysis can be caused by other organisms including protozoa and algae (Barak and Chet 1986). Synthetic substrates containing *p*-nitrophenyl (*p*-nitrophenyl glucoside, *p*-nitrophenyl sulfate, and *p*-nitrophenyl phosphate) are used as substrate by enzymes β-glucosidase, arylsulfatase, and phosphomonoesterase releasing *p*-nitrophenol after cleavage.

9.5.2.2 Substrate Utilization

Functional diversity of microbes can be determined using utilization of different substrates by microbes which include carbon, phosphorus, nitrogen, sulfur, etc. Soil microorganisms have capability to convert carbon into biomass. Carbon use efficiency determines the ratio of carbon used for respiration and biomass production. More biomass will be produced utilizing per unit of substrate if carbon use efficiency is high, which further leads to higher decomposition and more mineralization (Allison et al. 2010; Wieder et al. 2013). Along with carbon, presence of nitrogen also determines the rate of decomposition. If carbon nitrogen ratio is low, microbial communities will have higher carbon use efficiency leading to limitation of carbon. On the other hand, when C:N ratio will be high, excess of carbon will be lost in respiration as carbon use efficiency will be low as compared to available carbon.

Chemical composition of substrate also affects bioconversion process. Substrate having same C:N ratio but having different chemical structure converts into biomass with different efficiency. Complex substrate such as lignin needs role of different enzymes using different metabolic pathway, which determine loss of carbon in respiration (Manzoni et al. 2012). Soil profile also has impact on properties of substrate and availability of nutrients. During decomposition, C is lost as respiration, which leads to decrease in C:N ratio with increasing depth (Rumpel and Kögel-Knabner 2011). Slower microbial growth is observed with increase in substrate complexity as efficiency of microbial enzyme decreases. Under condition of nutrient distress, higher activity for oxidative enzymes were observed which reflect nutrient mining (Wild et al. 2015).

9.5.3 Cellular Constituents-Based Methods

The molecules that may create and preserve variations among lineages of microbes due to their conserved structures and differential functions are the ultimate determinants of diversity in microorganisms. Methods created to assess the connections between these molecules are therefore crucial for recognizing variety and classifying newly discovered species.

9.5.3.1 Nucleic Acids-Based Method

Deoxyribonucleic acid (DNA) and Ribonucleic acid (RNA) are two types of nucleic acids that are present in microbial cells. The most commonly used method for microbial identification using nucleic acids is polymerase chain reaction (PCR) as it is easy to perform, highly specific, sensitive, and has low detection limit. The most important and critical step in the analysis of samples for estimation of microbial diversity by PCR is nucleic acid extraction. Cell lysis is required to release the

nucleic acids which can be done by various physical and chemical methods, among which chemical lysis is most preferable.

9.5.3.2 Fluorescence In Situ Hybridization

Fluorescence in situ hybridization (FISH) is a nucleic acid-based method that involves the use of fluorescently labeled oligonucleotide probes which target a specific region of rRNA (16S/23S in Bacteria/Archaea) and is mainly used for the specific identification of microorganisms to the subspecies level in the medical and diagnostic field (Wagner and Haider 2012; Dias and Rathnayaka 2018; Oliveira et al. 2020). Fluorescent dyes primarily used for FISH are fluorescein, tetramethylrhodamine, and carbocyanine dyes, i.e., Cy3 and Cy5 (Rohde et al. 2015). The multicolor DOPE-FISH method uses probes that are double-labeled with different fluorophores at the 5′ and 3′ ends. The experiment showed that the new label combinations successfully detected six phylogenetically distant microbial taxa (Stoecker et al. 2010; Behnam et al. 2012). CLASI-FISH is a technique that allows for the detection of many taxa by labeling them with many different dyes and imaging them with spectral technology. The taxa are recognized by linear unmixing during postprocessing of the spectral pictures. CLASI-FISH has been effective in distinguishing up to 15 different target organisms in the same experiment (Valm et al. 2011; Valm et al. 2012; Mark Welch et al. 2016).

9.5.3.3 PCR-Based Method

The 16S rRNA molecule is an outstanding genetic tool for microbial community analysis due to its highly conserved nature among closely related species. The ~1500 bp 16S rRNA gene contains nine variable regions scattered through the extremely conserved 16S sequence and the sequencing of the entire gene was done via Sanger sequencing (Johnson et al. 2019). 16S rRNA sequence analysis is helpful for the diversity and phylogeny of microbes in any environment followed by bioinformatics analyses. PCR is used to amplify the desired 16S rRNA gene for further analyses. PCR mainly comprises three steps, i.e., denaturation, annealing, and extension. The first step is denaturation which is performed at 94 °C and is helpful in the separation of double-helix DNA. In the annealing step, the primers (short DNA oligonucleotides complementary to a specific sequence) bind to their corresponding sites on the template DNA, followed by the binding of DNA polymerase enzyme. The extension is the last step which is performed at 72 °C and the amplification of target DNA is completed in this step. All these three steps are repeated a number of times (a predetermined number of cycles) and millions of copies of the target DNA are formed. Reverse transcriptase-PCR (RT-PCR) is a method used to amplify complementary DNA (cDNA). Some public bioinformatics databases are accessible which are helpful for comparison of newly obtained

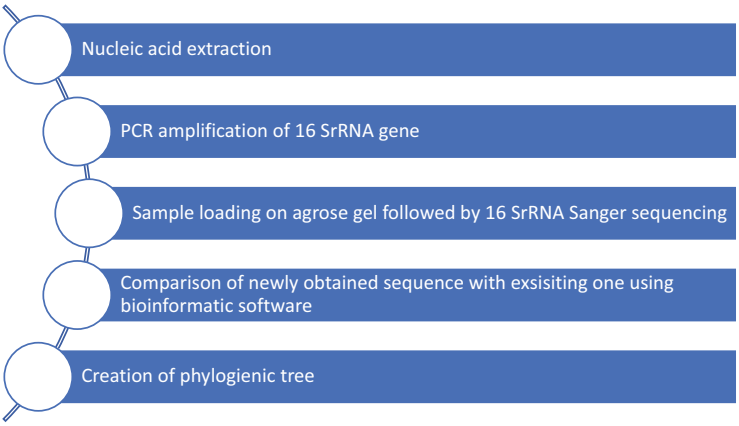

Fig. 9.2 Protocol for the identification of microbial diversity using nucleic acid

sequences to previously existing ones by performing a basic local alignment search tool (BLAST) followed by the creation of phylogenetic tree as depicted in Fig. 9.2.

However, 16S rRNA sequencing is not able to constantly identify bacteria up to the species level due to high sequence resemblances between some species (Deurenberg et al. 2017) like species of the *Streptococcus mitis* group cannot be differentiated by 16S rRNA sequencing due to 99–100% sequence similarities in genes (Lal et al. 2011). Similarly, *Escherichia coli* and *Shigella spp.* cannot be distinguished on the basis of the 16S rRNA genes due to >99% sequence identity (Devanga Ragupathi et al. 2017).

9.5.3.4 Next-Generation Sequencing

In a single sequence run, next-generation sequencing (NGS) controls the DNA sequence of the entire bacterial genome. NGS has many benefits over Sanger sequencing, including a greater resolution, a method independent of culture, accuracy in identifying microbial infections, etc. (Motro and Moran-Gilad 2017). A NGS method has been reported using PCR amplified products of the 16S–23S rRNA encoding region which showed high resolution and take less time for identification of bacteria compared to other methods (Benítez-Páez and Sanz 2017). However, this method has a few restrictions, including the lack of a comprehensive database for the 16S–23S rRNA encoding region and supplementary software that would have allowed for silent and accurate species identification. De novo assembly using CLC Genomics followed by BLAST-N using the NCBI database is the main technique which can be employed for identification of bacteria from clinical samples (Sabat et al. 2017).

9.5.3.5 PCR-Independent Methods

The PhyloChip technique has been established precisely for interrogating small subunit ribosomal RNA gene pools from multifaceted environmental samples (DeSantis et al. 2007; Hazen et al. 2010). The PhyloChip can distinguish between up to 8471 distinct taxa at the subfamily level, which is a substantial percentage of all known bacteria and archaea (Brodie et al. 2006). PhyloChip reproducibility and low cost help the investigators to analyze the microbial communities in varying environments. Two PCR-independent methods have been developed by DeAngelis et al. (2011) for direct evaluation of microbial communities, i.e., direct RNA hybridization along with double-strand cDNA generation as well as hybridization. Although direct hybridization of rRNA to oligonucleotide microarrays is not an innovative method (Call 2005) but this is the first implementation of 16S rRNA PhyloChip universally. In RNA hybridization technique, the half of the total RNA was 16S ribosomal RNA subsequently extra RNA was split and hybridized. Then, using the gene-specific primer 1492R, dscDNA was produced and put to each PhyloChip. The dscDNA was quantified, and then this DNA was broken up, tagged, and hybridized. rRNA was directly analyzed using a modified version of Cole et al. (2004).

9.5.4 Advance Methods

To understand the concept of microbial diversity and community composition of a sample, different advanced techniques have been developed by researchers and taxonomists like phospholipid fatty acid analysis (PLFA) and flow cytometry.

9.5.4.1 Phospholipid Fatty Acid Analysis

Cellular membranes of microbes chiefly comprise certain signature phospholipids which are swiftly converted to neutral lipids after the demise of the microorganism; therefore, PLFA analysis is a culture-independent method which characterizes living microbial community of particular location (Pinkart et al. 2002). Some PLFAs are features of specific functional groups of microbes which live in terrestrial and aquatic ecosystems. PLFA provides a balanced analysis of multifaceted microbial communities which cannot be differentiated by other cultivation-based methods and the evidence provided by PLFA is related to the phylogenetic study which is based on the 16S ribosomal RNA sequence homology (Findlay 1996). Distinct fatty acids are found in some taxa; hence it can be used as biomarker for the characterization of microbial groups. Most fatty acids do not exclusively belong to one type of organism. For instance, although cyclopropyl fatty acids are associated with anaerobic bacteria, they can also represent older gram-negative bacteria because they are capable of converting monoenoic PLFA 16:1:7c and 18:1:7c into cy17:0 and

cy19:0, respectively (Zelles 1999). Therefore, the ratios of the various fatty acids in the microorganisms are more dependable for their identification, particularly if they are present in the phospholipids.

The microbial lipid extract is segmented into neutral, glycol as well as phospholipids. Phospholipids are changed into fatty acid methyl esters by alkaline methanolysis. The final products can be analyzed by the Sherlock Microbial Identification System by comparing their holding times with bacterial acid standards. This system only identifies fatty acids with 9–20 carbons in acyl chain while longer fatty acids, i.e., up to 30 can be identified using gas chromatography along with mass spectrometry which can be helpful in phylogenetic lineages within microbial communities (Hu et al. 2006; Dickson et al. 2009). PLFA technique has been used for quantitative evaluations of the biomass of diverse taxonomic groups in samples and subsequently compared to the microbial taxa estimated from metabarcoding. They used two different sets of PLFA biomarkers having varied number of PLFAs to analyze the effect of PLFA biomarker selection on biomass estimates. Metabarcoding and PLFA analysis represented significantly different bacterial composition due to varied substrates. They showed the most prominent differences for the gram-negative bacteria, which were overestimated by metabarcoding compared to PLFA analysis (Lewe et al. 2021).

9.5.4.2 Flow Cytometry

Flow cytometry is a culture-independent, powerful, swift, and very sensitive technique for monitoring, enumeration as well as characterization of microorganisms via quantitative count of microbial cells. It also gives information regarding physiological and structural characteristics of microbes as well as their viability (Kennedy and Wilkinson 2017; Rajapaksha et al. 2019). The technique of flow cytometry is very useful for the identification and monitoring of microbial contamination in water, air, and on biotic and abiotic surfaces (such as conveyor belts or pipelines) (i.e., solid food samples). The flow cytometer primarily consists of a flow chamber, a light source (such as a laser or mercury lamp), dichroic mirrors for beam focusing, bandpass filters for identifying different wavelengths, detectors (such as photodiodes and photomultiplier tubes for monitoring and amplification of the signals), and data processing (Paparella et al. 2012; Wu et al. 2016).

Props et al. (2016) established a flow cytometry-based computational technique to monitor the microbial biodiversity with high resolution. The biodiversity of microbes can be swiftly measured from community, single cell as well as phenotypic data (morphology and nucleic acid content) using above method. The yield of this developed computational method is that the phenotypic diversity of microbes is strongly associated to the species diversity, using regressions and correlation analysis as anticipated by 16S rRNA gene sequencing. OzelDuygan et al. (2020) created the CellCognize supervised machine learning method to recognize standard cell types and monitor sample variety using flow cytometric characteristics of individual cells in comparison to strain and bead standards. With 32 microorganism cell and

bead standards, they were able to create neural networks which resulted in classifiers that were then applied to flow cytometric data to estimate the cell types of untrained microbiological samples with known or unknown composition. The resulting classifiers had an average prediction accuracy of 80% after rigorous in silico validation on known bacteria.

9.6 Conclusion

The microbial population represents the richest biodiversity resources on the planet earth. The utilization and development of microbial resources are crucial for resource exploitation, conservation of global genetic resources, exploration of biotechnologically valuable microorganisms, monitoring and prediction of environmental changes by studying the diversity patterns, and for the assessment of the impact of microorganisms on the maintenance of human health. The importance of microbial diversity research makes it one of the primary driving forces in the advancement of life science in the twenty-first century. Recent advances in genetic technologies have accelerated microbial diversity research by allowing for high-throughput screening and analysis in conjunction with powerful bioinformatics platforms.

References

Allison SD, Wallenstein MD, Bradford MA (2010) Soil-carbon response to warming dependent on microbial physiology. Nat Geosci 3:336–340. https://doi.org/10.1038/ngeo846

Barak R, Chet I (1986) Determination, by fluorescein diacetate staining, of fungal viability during mycoparasitism. Soil Biol Biochem 18(3):315–319

Behnam F, Vilcinskas A, Wagner M, Stoeckerb K (2012) A straightforward DOPE (double labeling of oligonucleotide probes)-FISH (fluorescence in situ hybridization) method for simultaneous multicolor detection of six microbial populations. Appl Environ Microbiol 78:5138–5142. https://doi.org/10.1128/AEM.00977-12

Benítez-Páez A, Sanz Y (2017) Multi-locus and long amplicon sequencing approach to study microbial diversity at species level using the MinIONTM portable nanopore sequencer. Gigascience 6:1–12. https://doi.org/10.1093/gigascience/gix043

Bernard L, Mougel C, Maron PA, Nowak V, Lévêque J, Henault C, Ranjard L (2007) Dynamics and identification of soil microbial populations actively assimilating carbon from 13C-labelled wheat residue as estimated by DNA-and RNA-SIP techniques. Environ Microbiol 9(3):752–764

Besset-Manzoni Y, Rieusset L, Joly P, Comte G, Prigent-Combaret C (2018) Exploiting rhizosphere microbial cooperation for developing sustainable agriculture strategies. Environ Sci Pollut Res 25(30):29953–29970

Brodie EL, DeSantis TZ, Joyner DC, Baek SM, Larsen JT, Andersen GL et al (2006) Application of a highdensity oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. Appl Environ Microbiol 72:6288–6298

Brown JW (2014) What is microbial diversity. In: Brown JW (ed) Principles of microbial diversity. ASM Press, Washington DC, pp 5–14

Busse MD, Sanchez FG, Ratcliff AW, Butnor JR, Carter EA, Powers RF (2009) Soil carbon sequestration and changes in fungal and bacterial biomass following incorporation of forest residues. Soil Biol Biochem 41(2):220–227

Call DR (2005) Challenges and opportunities for pathogen detection using DNA microarrays. Crit Rev Microbiol 31:91–99

Chimileski S, Kolter R (2017) Life at the edge of sight: a photographic exploration of the microbial world. Harvard University Press, Cambridge

Cole K, Truong V, Barone D, McGall G (2004) Direct labeling of RNA with multiple biotins allows sensitive expression profiling of acute leukemia class predictor genes. Nucleic Acids Res 32(11):e86

Constant P, Chowdhury SP, Pratscher J, Conrad R (2010) Streptomycetes contributing to atmospheric molecular hydrogen soil uptake are widespread and encode a putative high-affinity [NiFe]-hydrogenase. Environ Microbiol 12:821–829

Constant P, Poissant L, Villemur R (2008) Isolation of *Streptomyces* sp. PCB7, the first microorganism demonstrating high-affinity uptake of tropospheric H2. ISME J 2:1066–1076

DeAngelis KM, Wu CH, Beller HR, Brodie EL, Chakraborty R, DeSantis TZ, Fortney JL, Hazen TC, Osman SR, Singer ME, Tom LM, Andersen GL (2011) PCR amplification-independent methods for detection of microbial communities by the high-density microarray PhyloChip ABSTRACT. Appl Environ Microbiol 77(18):6313–6322. https://doi.org/10.1128/AEM.05262-11

DeLong DC (1996) Defining biodiversity, vol 24. Wildlife Society Bulletin, pp 738–749

DeSantis TZ, Brodie EL, Moberg JP, Zubieta IX, Piceno YM, Andersen GL (2007) High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. Microb Ecol 53:371–383

Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, García-Cobos S et al (2017) Application of next-generation sequencing in clinical microbiology and infection prevention. J Biotechnol 243:16–24. https://doi.org/10.1016/j.jbiotec.2016.12.022

Devanga Ragupathi NK, Muthuirulandi Sethuvel DP, Inbanathan FY, Veeraraghavan B (2017) Accurate differentiation of Escherichia coli and Shigella serogroups: challenges and strategies. New Microbes New Infect 21:58–62. https://doi.org/10.1016/j.nmni.2017.09.003

Dias G, Rathnayaka U (2018) Fluorescence in situ hybridization (FISH) in food pathogen detection. Int J Mol Sci 3:143–149

Dickson L, Bull ID, Gates PJ, Evershed RP (2009) A simple modification of a silicic acid lipid fractionation protocol to eliminate free fatty acids from glycolipid and phospholipid fractions. J Microbiol Methods 78:249–254

Ducklow H (2008) Microbial services: challenges for microbial ecologists in a changing world. Aquat Microb Ecol 53(1):13–19

Dunfield PF, Conrad R (2000) Starvation alters the apparent half-saturation constant for methane in the type II methanotroph Methylocystis strain LR1. Appl Environ Microbiol 66:4136–4138

Fenchel T (2002) Microbial behavior in a heterogeneous world. Science(New York, NY) 296(5570):1068–1071. https://doi.org/10.1126/science.1070118

Findlay R (1996) The use of phospholipid fatty acids to determine microbial community structure. In: Akkermans ADL, van Elsas JD, de Bruijn F (eds) Molecular microbial ecology manual. Kluwer academic publishers, Dordrecht, The Netherlands, pp 1–17

Giller KE, Witter E, McGrath SP (2009) Heavy metals and soil microbes. Soil Biol Biochem 41(10):2031–2037

Hashimoto T, Hattori T (1989) Grouping of soil bacteria by analysis of colony formation on agar plates. Biol Fertil Soils 7:198–201

Hazen TC, Dubinsky EA, DeSantis TZ, Andersen GL, Piceno YM, Singh N, Jansson JK, Probst A, Borglin SE, Fortney JL et al (2010) Deep-sea oil plume enriches indigenous oil-degrading Bacteria. Science 330:204–208. https://doi.org/10.1016/j.mimet.2021.106271

Heaton BE, Herrou J, Blackwell AE, Wysocki VH, Crosson S (2012) Molecular structure and function of the novel BrnT/BrnA toxin-antitoxin system of Brucella abortus. J Biol Chem 287(15):12098–12110

Hu J, Zhang H, Peng P (2006) Fatty acid composition of surface sediments in the subtropical Pearl River estuary and adjacent shelf, southern China. Estuar Coast Shelf Sci 66:346–356

Islam ST, Vergara Alvarez I, Saïdi F, Guiseppi A, Vinogradov E, Sharma G, Espinosa L, Morrone C, Brasseur G, Guillemot JF et al (2020) Modulation of bacterial multicellularity via spatio-specific polysaccharide secretion. PLoS Biol 18(6):e3000728. https://doi.org/10.1371/journal.pbio.3000728

Johnson JS, Spakowicz DJ, Hong BY, Petersen LM, Demkowicz P, Chen L, Leopold SR, Hanson BM, Agresta HO, Gerstein M et al (2019) Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat Commun 10:5029. https://doi.org/10.1038/s41467-019-13036-1

Kasahara Y, Hattori T (1991) Analysis of bacterial populations in a grassland soil according to rates of development on solid media. FEMS Microbiol Ecol 86:95–102

Kennedy D, Wilkinson MG (2017) Application of flow cytometry to the detection of pathogenic bacteria. Curr Issues Mol Biol 23:21–38

Koskella B, Meaden S, Crowther WJ, Leimu R, Metcalf CJE (2017) A signature of tree health? Shifts in the microbiome and the ecological drivers of horse chestnut bleeding canker disease. New Phytol 215(2):737–746

Lal D, Verma M, Lal R (2011) Exploring internal features of 16S rRNA gene for identification of clinically relevant species of the genus streptococcus. Ann Clin Microbiol Antimicrob 10:28. https://doi.org/10.1186/1476-0711-10-28

Lewe N, Hermans S, Lear G, Kelly LT, Thomson-Laing G, Weisbrod B, Wood SA (2021) Phospholipid fatty acid (PLFA) analysis as a tool to estimate absolute abundances from compositional 16S rRNA bacterial metabarcoding data. J Microbiol Methods 188:106271

Liu WT, Marsh TL, Cheng H, Forney LJ (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. Appl Environ Microbiol 63(11):4516–4522. https://doi.org/10.1128/aem.63.11.4516-4522.1997

Luna GM, Dell'Anno A, Giuliano L, Danovaro R (2004) Bacterial diversity in deep Mediterranean sediments: relationship with the active bacterial fraction and substrate availability. Environ Microbiol 6(7):745–753

Maiden MC, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND (2013) MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Microbiol 11(10):728–736. https://doi.org/10.1038/nrmicro3093

Manzoni S, Taylor P, Richter A, Porporato A, Ågren GI (2012) Environmental and stoichiometric controls on microbial carbon-use efficiency in soils. New Phytol 196:79–91

Maraha N (2007) Physiological status of bacteria used for environmental applications. Karolinska Institutet (Sweden)

Maraha N, Backman A, Jansson JK (2004) Monitoring physiological status of GFP-tagged *Pseudomonas fluorescens* SBW25 under different nutrient conditions and in soil by flow cytometry. FEMS Microbiol Ecol 51(1):123–132

Mark Welch JL, Rossetti BJ, Rieken CW, Dewhirst FE, Borisy GG (2016) Biogeography of a human oral microbiome at the micron scale. Proc Natl Acad Sci U S A 113:E791–E800. https://doi.org/10.1073/pnas.1522149113

Matulich KL, Mohamed DJ, Martiny JBH (2013) Microbial biodiversity. In: Encyclopedia of biodiversity. Elsevier, pp 252–258

McFall-Ngai MJ (2015) Giving microbes their due–animal life in a microbially dominant world. J Exp Biol 218(12):1968–1973

Motro Y, Moran-Gilad J (2017) Next-generation sequencing applications in clinical bacteriology. Biomol Detect Quantif 14:1–6. https://doi.org/10.1016/j.bdq.2017.10.002

Naranjo-Ortiz MA, Gabaldón T (2019) Fungal evolution: diversity, taxonomy and phylogeny of the fungi. Biol Rev Camb Philos Soc 94(6):2101–2137. https://doi.org/10.1111/brv.12550

Nèble S, Calvert V, Le Petit J, Criquet S (2007) Dynamics of phosphatase activities in a cork oak litter (*Quercus suber* L.) following sewage sludge application. Soil Biol Biochem 39(11): 2735–2742

Oliveira R, Almeida C, Azevedo NF (2020) Detection of microorganisms by fluorescence in situ hybridization using peptide nucleic acid. In: Nielsen PE (ed) Peptide nucleic acids: methods and protocols. springer, New York, NY, pp 217–230

Oliver SP, Jayarao BM, Almeida RA (2005) Foodborne pathogens in milk and the dairy farm environment: food safety and public health implications. Foodborne Pathog Dis 2(2):115–129

OzelDuygan BD, Hadadi N, Babu AF, Seyfried M, Meer JRV (2020) Rapid detection of microbiota cell type diversity using machine-learned classification of flow cytometry data. Commun Biol 3: 379. https://doi.org/10.1038/s42003-020-1106-y

Paparella A, Serio A, Chaves C (2012) Flow cytometry applications in food safety studies. In: Flow Cytometry - Recent Perspectives, pp 69–102. https://doi.org/10.5772/37904

Pinkart HG, Ringelberg DB, Piceno YM, MacNaughton SJ, White DC (2002) Biochemical approaches to biomass measurements and community structure. In: Hurst CJ, Crawford RL, Knudsen GR, McInerney MJ, Stetzenbach LD (eds) Manual of environmental microbiology. American Society for Microbiology press, Washington, DC, pp 101–113

Props R, Monsieurs P, Mysara M, Clement L, Boon N (2016) Measuring the biodiversity of microbial communities by flow cytometry. Methods Ecol Evol 7:1376–1385. https://doi.org/10.1111/2041-210X.12607

Purvis A, Hector A (2000) Getting the measure of biodiversity. Nature 405(6783):212–219. https://doi.org/10.1038/35012221

Rajapaksha P, Elbourne A, Gangadoo S, Brown R, Cozzolino D, Chapman JA (2019) Review of methods for the detection of pathogenic microorganisms. Analyst 144:396–411

Rohde A, Hammerl JA, Appel B, Dieckmann R, Al Dahouk S (2015) FISHing for bacteria in food—a promising tool for the reliable detection of pathogenic bacteria? Food Microbiol 46: 395–407

Roszak DB, Colwell R (1987) Survival strategies of bacteria in the natural environment. Microbiol Rev 51(3):365–379

Rumpel C, Kögel-Knabner I (2011) Deep soil organic matter—a key but poorly understood component of terrestrial C cycle. Plant Soil 338(1):143–158

Sabat AJ, Van Zanten E, Akkerboom V, Wisselink G, Van Slochteren K, De Boer RF et al (2017) Targeted next-generation sequencing of the 16S-23S rRNA region for culture-independent bacterial identification-increased discrimination of closely related species. Sci Rep 7:3434. https://doi.org/10.1038/s41598-017-03458-6

Shoemaker WR, Locey KJ, Lennon JT (2017) A macroecological theory of microbial biodiversity. Nat Ecol Evol 1(5):107. https://doi.org/10.1038/s41559-017-0107

Stiehl-Braun PA, Powlson DS, Poulton PR, Niklaus PA (2011) Effects of N fertilizers and liming on the micro-scale distribution of soil methane assimilation in the long-term park grass experiment at Rothamsted. Soil Biol Biochem 43(5):1034–1041

Stoecker K, Dorninger C, Daims H, Wagner M (2010) Double labeling of oligonucleotide probes for fluorescence in situ hybridization (DOPE-FISH) improves signal intensity and increases rRNA accessibility. Appl Environ Microbiol 76:922–926. https://doi.org/10.1128/AEM.02456-09

Swingland I (2013) Biodiversity definition. In: Levin SA (ed) Encyclopedia of biodiversity, 2nd edn. Academic Press, Amsterdam, The Netherlands, pp 399–410

Tang SKY, Marshall WF (2018) Cell learning. Curr Biol 28(20):R1180–R1184. https://doi.org/10.1016/j.cub.2018.09.015

Valm AM, Welch JLM, Rieken CW, Hasegawa Y, Sogin ML, Oldenbourg R et al (2011) Systems-level analysis of microbial community organization through combinatorial labeling and spectral imaging. Proc Natl Acad Sci U S A 108:4152–4157. https://doi.org/10.1073/pnas.1101134108

Valm AM, Mark Welch JL, Borisy GG (2012) CLASI-FISH: principles of combinatorial labeling and spectral imaging. Syst Appl Microbiol 35:496–502. https://doi.org/10.1016/j.syapm.2012.03.004

van Teeseling MCF, de Pedro MA, Cava F (2017) Determinants of bacterial morphology: from fundamentals to possibilities for antimicrobial targeting. Front Microbiol 8:1264

Wagner M, Haider S (2012) New trends in fluorescence in situ hybridization for identification and functional analyses of microbes. Curr Opin Biotechnol 23:96–102

Wieder WR, Bonan GB, Allison SD (2013) Global soil carbon projections are improved by modelling microbial processes. Nat Clim Chang 3(10):909–912

Wild B, Schnecker J, Knoltsch A, Takriti M, Mooshammer M, Gentsch N, Mikutta R, Alves RJE, Gittel A, Lashchinskiy N, Richter A (2015) Microbial nitrogen dynamics in organic and mineral soil horizons along a latitudinal transect in western Siberia. Glob Biogeochem Cycles 29:567–582

Winding A, Hund-Rinke K, Rutgers M (2005) The use of microorganisms in ecological soil classification and assessment concepts. Ecotoxicol Environ Saf 62(2):230–248

Wu LN, Wang S, Song YY, Wang X, Yan XM (2016) Applications and challenges for single-bacteria analysis by flow cytometry. Sci China Chem 59:30–39

Yang CY, Bialecka-Fornal M, Weatherwax C, Larkin JW, Prindle A, Liu J, Garcia-Ojalvo J, Süel GM (2020) Encoding membrane-potential-based memory within a microbial community. Cell Systems 10(5):417–423. e3. https://doi.org/10.1016/j.cels.2020.04.002

Zelles L (1999) Identification of single cultured micro-organisms based on their whole-community fatty acid profiles, using an extended extraction procedure. Chemosphere 39:665–682

Chapter 10
Microbial Diversity and Its Relevance to Animal Health

Geetika Kaur, Nikhlesh K. Singh, and Geetika Verma

Abstract Microbial diversity is defined as a vast variety of unicellular microorganisms. Microbes are involved in a wide array of metabolic activities in diverse and extreme environments; thus, creating a microenvironment for the survival and evolution of other living organisms. From an ecological perspective, it is interesting to note that microbial diversity is supported by a variety of selection theories, mutations, and environmental heterogeneity. Microbial species-inhabited animal microbiome is essential for providing nutritional input, maintaining health, control of pathogens, growth, reproduction, and preventing diseases. Advancements in microbiological techniques and microbial coinhabitants have paved a way to enhance animal health. This chapter provides an understanding on the scope of microbial diversity and its relevance to animal health.

Keywords Microbial diversity · Animal health · Microorganisms · Selection theories · Microbiome

10.1 Introduction to Microbial Diversity

Microorganisms are predominantly smaller life forms that are enormous in diversity. Microbial diversity is defined as a vast variety of unicellular microorganisms. It includes bacteria, archaea, protists, and fungi (Dunlap 2001). All these organisms have distinguishable characteristics in terms of morphology, physiology, genomic structure, and cellular metabolism. The term mycobiome refers to the fungal

G. Kaur (✉) · N. K. Singh
Integrative Biosciences Center, Wayne State University, Detroit, MI, USA

Department of Ophthalmology, Visual and Anatomical Sciences, School of Medicine, Wayne State University, Detroit, MI, USA
e-mail: geetika.kaur@wayne.edu; nsingh2@wayne.edu

G. Verma
Department of Experimental Medicine and Biotechnology, Post Graduate Institute of Medical Education and Research, Chandigarh, India

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_10

microbiome and plays a crucial role in animal health along with bacteriome (Summers and Arfken 2022). In the past few decades, microbes, their diversification, and their role in different aspects of the ecosystem have gained biological importance. They are involved in a wide array of metabolic activities in diverse and extreme environments; thus, creating a microenvironment for the survival and evolution of other living organisms (Rampelotto 2013). Moreover, microorganisms perform essential functions for the development of the ecosystem and from a health perspective. The diversified microbial forms not only contribute to human health but are associated with animal health and diseases. Therefore, understanding and exploring microbial diversity has become indispensable.

10.2 Scope and Significance of Microbial Diversity

The planet earth is known as a "microbial planet" in the "Age of Bacteria" due to the existence of microbes for more than 3.8 billion years on earth. Scientific evidence indicates that these microbes account for the major kind of organisms that have continued to occupy all the existing niches on earth (Dunlap 2001; Rampelotto 2013). The true extent of diversity of microorganisms is unspecified and their estimations are mere guesses. Microbiologists are aware of the fact that many microbes exist in nature and their new diversified forms are yet to be explored and discovered. With the recent advances in science and technology, the discovery of new microbial species is gaining momentum (Dunlap 2001). From an ecological perspective, it is interesting to note that microbial diversity is supported by a variety of selection theories, mutations, and environmental heterogeneity (Kneital 2008; Hibbing et al. 2010; Thursby and Juge 2017; Raynes et al. 2018).

10.2.1 Niche Exclusion Principle

According to the niche exclusion principle, only a single kind of species or genotype is supported in a particular or single niche or patch. The diversity in the microbial forms depends on the variety of niches available. Thus, microbes are considered as the driving forces as they can grow and adapt in varied environmental conditions such as aerobic and anaerobic conditions, low and high oxygen levels, pH, temperature, and pressure. The differential microbes, their actions, and interactions with the environment form new niches, leading to more genetic diversification in microbial world (Hibbing et al. 2010).

10.2.2 Mutation

Secondly, mutation plays a crucial role in genetic diversification. Mutation within the species results in the formation or survival of new and different genetic variants. These mutations are deleterious or beneficial for the survival and growth of the environment and this process is documented as periodic selection (Raynes et al. 2018). The concept of periodic selection contradicts the niche exclusion principle due to the evolution of molecular varied species in a single niche. The number of mutations that occur in a population are directly proportional to the number of genetic variants (Raynes et al. 2018).

10.2.3 Environmental Heterogeneity

With the evolutionary trends and development of multicellular organisms such as plants and animals, more diversified and widely distributed microbial colonies have been observed due to a complex environment. The complex or heterogeneous environment is composed of different types of niches. The growth and colonization of differential microbes are highly favored by the niches provided by these multicellular organisms. For instance, the human body provides a multitude of habitats that are physiochemically distinct such as skin, mucous membrane, and gastrointestinal system (Thursby and Juge 2017). Specific microbes establish a symbiotic relationship with other organisms for example, *Rhizobium*, a nitrogen-fixing bacteria with leguminous plants, and *Vibrio*, a bioluminescent with animals including fishes and cephalopods (Thursby and Juge 2017; Raynes et al. 2018).

10.3 Impact on Animal Health

Microbial species inhabiting animal microbiome is essential for providing nutritional input, maintaining health, control of pathogens, growth, reproduction, and preventing diseases. The gastrointestinal microbiomes are the best examples, which help in the conversion of food and detoxification of certain plant materials in ruminants (Rowland et al. 2018). Animal studies have shown that the major phyla inhabiting gastrointestinal tracts of animals form a stable population of gram-negative population, *Proteobacteria* and *Bacteroidetes* and gram-positive population of *Firmicutes* (*Clostridiales* and *Lactobacillales*). However, the microbiome differs between the animals, based on their eating habits, for example, carnivores, omnivores, and herbivores. It is important to mention here that any disruption or change in microbiota or their distribution may lead to microbial imbalance or "dysbiosis" (Patel 2021).

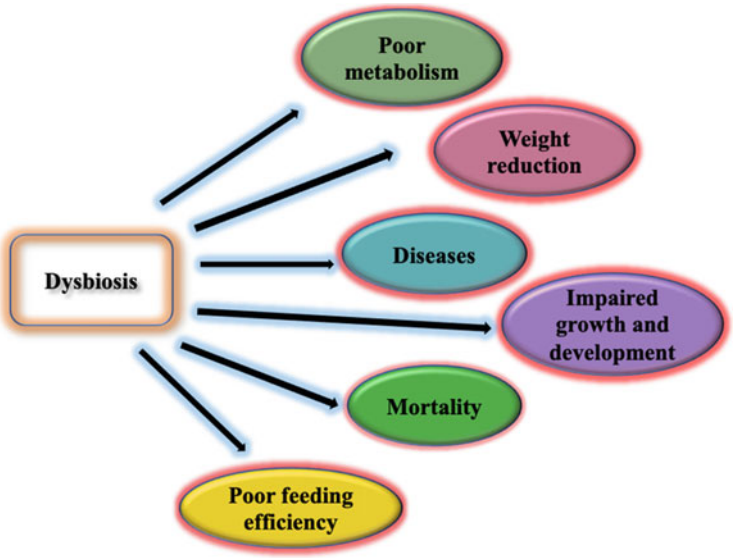

Fig. 10.1 Impact of dysbiosis on animal health

The host–microbiome association in animals largely depends on the niche, i.e., internal and external environments. The external environment is modulated by temperature, pressure, and precipitation; whereas the internal environment is regulated by the host factors, like the immune system, diet, and phylogeny. Both the environmental factors influence the type, growth, and functioning of the microbial forms in a particular biome (Belkaid and Hand 2014). The number of pathogens that are responsible for the development of the disease may surge due to variations in the microbiome.

With the emergence of sophisticated techniques such as high-throughput DNA-based sequencing along with bioinformatic tools, the characterization of microbes in the animal niche has been possible. The microbes in or on the body of animals may enhance or hinder the growth, metabolism, and overall health of the host (Fig. 10.1).

The effects of microbiota on different animal types include:

10.3.1 Weight

Ruminants such as bovines and cattle are a portable source of milk, leather for clothing, and energy to tile in farms and fields. These animals rely on microbial fermentation for nutrient and energy requirements. Rumen microbiome consists of more than 30 different phyla and 90 genera, which enhance their digestion that helps in significant weight gain to improve meat production. Microbes such as *Firmicutes*,

Proteobacteria, *Actinobacteria*, and *Bacteroidetes* form the microbiome that plays an essential role in efficient nutrition uptake to produce high-quality milk and meat. Dysbiosis in ruminants may lead to reduced weight and thereby hinder the growth and development of the animal (De Freitas et al. 2020).

10.3.2 Growth and Development

Poultry is a major source of proteins for humans. Age, diet, microbial diversity, and antibiotics impact the poultry gut and thus, affecting the growth, development, and life-cycle of the bird. The gut microbiome of the bird comprises of 15 phyla and 288 genera. Amongst these 15 phyla, *Firmicutes* and *Bacteroidetes* are the dominant ones. Antibiotics such as enrofloxacin and diclazuril are used to reduce pathogens and increase the growth yield. However, these antibiotic usages have long-term effects on the gut microbiota and influence the number of *Lactobacillus, Streptococcus,* and *Clostridium* (Xiao et al. 2017).

10.3.3 Feeding Efficiency

The porcine microbiome not only focuses on the microbes inhabited in the pig gut but also their role in maintaining health and production. Pigs are well-suited models to study human systems and diseases as they share anatomical similarities. Xiao et al. (2016) studied the pig fecal microbiome that influences the health, sex, age, growth, fatness, and genetics of the animal. *Clostridium*, *Lactobacillus*, *Eubacterium,* and *Prevotella* are among the predominant species present in the fecal microbiome. The author suggested that extensive diversity is still undiscovered in the pig microbiome. The feeding efficiency is associated with gut microbiome in pigs. Microbial families such as *Lachnospiraceae* and *Prevotellaceae* and the genera included *Escherichia, Streptococcus,* and *Shigella* are linked with low feeding efficiency; whereas *Streptococcus gallolyticus* is potentially associated with improved feeding efficiency (Xiao et al. 2016).

10.3.4 Microbe-Associated Diseases

Dogs, cats, and equines are amongst the most common and popular companion animals. In these animals, the oral and gut microbiome is related to their nutrition uptake and health. The most common phyla inhabiting the animal microbiome are *Fusobacteria, Firmicutes, Actinobacteria, Proteobacteria,* and *Bacteroidetes.* In these companion animals, oral disease such as periodontal disease is the commonest problem that occurs due to a decrease in health-associated microbiomes, for

example, *Bergeyella zoohelcum*, *Moraxella* spp., *Neisseria shayeganii*, and *Capnocytophaga* (Ruparell et al. 2020). The loss or disruption in the commensal microbiota often leads to the changes in the metabolites of the bacteria like secondary bile acids and short chain fatty acids which ultimately leads to acute or chronic intestinal inflammation in dogs and cats (Suchodolski 2016).

10.3.5 Mortality

Marine animals include fish, crabs, eels, shrimp, prawns, molluscs, and others. In the gut of marine animals, more than 30 phyla are present with the highest amounts of *Bacteroidetes* and *Proteobacteria*. *Vibrio* is a bacterial contaminant that causes devastating effects on the mortality rate. Therefore, antibiotics have been used to provide protection against contaminants to produce quality food. Antibiotic usage is now replaced by a sustainable method of introducing beneficial bacteria or probiotics to improve growth and development, for example, *Endozoicomonas* (de Bruijn et al. 2018).

Dysbiosis in the microbiome is considered a serious threat to the wildlife population. It causes an overall decline in the wildlife population itself or their health. The proliferation of zoonoses causes a threat to wildlife including bats, rodents, and primates. For instance, fungal pathogens, *B. dendrobatidis, Batrachochytrium salamandrivorans, and Ophidiomyces ophiodiicola* are associated with the decline in amphibian and reptile populations; whereas *Pseudogymnoascus destructans*, the psychrophilic fungus is associated with mortality in the bat population (Williams et al. 2018).

Furthermore, animals harbor a wide variety of bacteria and viruses that cause diseases in humans and other animals. Recently, coronavirus disease (COVID-19) has emerged as a global threat and pandemic. For this, bat species are speculated to be the reservoir of coronavirus. Evidence in the literature reported that bat harbors more than 200 coronavirus species that cause mortality in humans as well as animals of agriculture importance (Banerjee et al. 2019).

10.4 Microbial-Based Therapies

The microbial forms, humans, animals, and the environment share a common relationship, forming one health system. All these factors are equally important and play a vital role in maintaining the ecosystem. This approach is required to study the pathogenic and nonpathogenic microbes and their mode of transfer and transmission among each other. Advancements in microbiological techniques and microbial coinhabitants have paved a way to improve animal health via:

1. Qualitative and quantitative production of animal products including meat, eggs, milk, and fiber using prebiotics, probiotics, and synbiotics
2. Protection against pathogens and disease mitigation
3. Dietary changes affecting animal microbiomes (Mimee et al. 2016)

The health benefits may be attained by using microbial-based therapies, which include:

10.4.1 Dietary Changes

A proper and healthy diet also permits the growth of beneficial bacteria. The animal microbiome can be maintained by making necessary dietary changes. The introduction of prebiotics and probiotics has revolutionized the microbiome world. Prebiotics include fructo-oligosaccharides, inulin, and high-fiber diets; and commercial probiotics include Lactobacilli and *Bifidobacteria.* The use of pre- and probiotics helps in the reduction of potential pathogens such as *C. perfringens* and *E. coli* (Mimee et al. 2016; Gilbert and Stephens 2018). Further, a combination of prebiotics and probiotics, known as synbiotics, has also been employed in improving animal health and performance. Many studies have shown that the use of synbiotics can diminish inflammation, bad cholesterol levels, remove toxins, activate the metabolism, and promote the growth of microorganisms beneficial to the host (Markowiak and Śliżewska 2018).

10.4.2 Symbiotic Relationship Management and Microbiome Manipulation

Dysbiosis is one of the major problems that interfere with the host–microbe relationship and affect the host's health. Thus, targeting this process can provide beneficial results to improve and promote animal health by governing and regulating the immune system. Transmission of symbionts, partner fidelity, and host-bacteria association play a vibrant role in this process. For instance, mucus in the intestinal system is a prominent site for the transmission of symbiont bacteria that helps in pathogen control, endocrine signaling, and nutrition improvement (Gilbert and Stephens 2018).

Microbiome manipulation has been primarily used for the protection and remediation of coral reef areas. This can be attained by directly introducing or bioaugmenting specific microbial strains. Laboratory and field studies in the literature proved that the concept of microbiome manipulation is possible and can protect the host against environmental threats. Reports have shown that manipulation of mycobiome, over bacteriome, by diet or other environmental factors serves as a potential candidate for dietary interventions in promoting animal health and

performance (Summers and Arfken 2022). The process of microbiome manipulation includes a simple technique of identification of beneficial microorganisms followed by increasing their abundance to reduce dysbiosis and maintaining a symbiotic relationship. Microbiota in an unhealthy animal can be maintained by transferring intestinal contents of healthy animals to it after characterization of specific bacteria, for example, *C. difficile* infection is a cause of concern. Intestinal content with specific bacteria that can constrain the growth of *C. difficile* can be used for transfaunation in animals (Kashyap et al. 2017).

10.4.3 Customized Microbial-Based Therapies

The framework of customized microbial-based therapies is proposed and certainly specific for the target animal species. It involves a combination of (a) manipulating the internal environment via prebiotics, diet, and antimicrobials, depending on the microbiome studies; and (b) animal rotation that focuses on beneficial microbes that suppress the growth of pathogens. The specific combination of customized therapy may provide similar benefits in different animals (Banerjee and Ray 2017).

10.5 Conclusion

The knowledge of animal host–microbial dynamics is not only helpful in improving animal health but also provides us insights into the use of microbial therapies for human health. Microbial therapies can promote human health by manipulating specific microbes and forming a symbiotic relationship in the host. Microbial diversity is very vast and versatile. Further, study and discovery of new microbiota will help to comprehend the role of different microorganisms in various diseases and designing new microbial therapies against them.

References

Banerjee G, Ray AK (2017) The advancement of probiotics research and its application in fish farming industries. Res Vet Sci 115:66–77. https://doi.org/10.1016/j.rvsc.2017.01.016

Banerjee A, Kulcsar K, Misra V, Frieman M, Mossman K (2019) Bats and Coronaviruses. Viruses 11(1):41. https://doi.org/10.3390/v11010041

Belkaid Y, Hand TW (2014) Role of the microbiota in immunity and inflammation. Cell 157(1): 121–141. https://doi.org/10.1016/j.cell.2014.03.011

de Bruijn I, Liu Y, Wiegertjes GF, Raaijmakers JM (2018) Exploring fish microbial communities to mitigate emerging diseases in aquaculture. FEMS Microbiol Ecol 94(1):161. https://doi.org/10.1093/femsec/fix161

De Freitas AS, de David DB, Takagaki BM, Roesch LFW (2020) Microbial patterns in rumen are associated with gain of weight in beef cattle. Antonie Van Leeuwenhoek 113(9):1299–1312. https://doi.org/10.1007/s10482-020-01440-3

Dunlap PV (2001) Microbial diversity. In: Levin SA (ed) Encyclopedia of biodiversity, 2nd edn. Academic Press, pp 280–291

Gilbert JA, Stephens B (2018) Microbiology of the built environment. Nat Rev Microbiol 16(11): 661–670. https://doi.org/10.1038/s41579-018-0065-5

Hibbing ME, Fuqua C, Parsek MR, Peterson SB (2010) Bacterial competition: surviving and thriving in the microbial jungle. Nat Rev Microbiol 8(1):15–25. https://doi.org/10.1038/nrmicro2259

Kashyap PC, Chia N, Nelson H, Segal E, Elinav E (2017) Microbiome at the frontier of personalized medicine. Mayo Clin Proc 92(12):1855–1864. https://doi.org/10.1016/j.mayocp.2017.10.004

Kneital J (2008) Gause's competitive exclusion principle. In: Jørgensen SE, Fath BD (eds) Encyclopedia of ecology. Academic press, pp 1731–1734

Markowiak P, Śliżewska K (2018) The role of probiotics, prebiotics and synbiotics in animal nutrition. Gut pathogens 10:21. https://doi.org/10.1186/s13099-018-0250-0

Mimee M, Citorik RJ, Lu TK (2016) Microbiome therapeutics—advances and challenges. Adv Drug Deliv Rev 105:44–54. https://doi.org/10.1016/j.addr.2016.04.032

Patel V (2021) The gut microbiome. In: Short E (ed) A prescription for healthy living. Academic press, pp 165–175. https://doi.org/10.1016/B978-0-12-821573-9.00015-1

Rampelotto PH (2013) Extremophiles and extreme environments. Life (Basel, Switzerland) 3(3): 482–485. https://doi.org/10.3390/life3030482

Raynes Y, Wylie CS, Sniegowski PD, Weinreich DM (2018) Sign of selection on mutation rate modifiers depends on population size. Proc Natl Acad Sci U S A 115(13):3422–3427. https://doi.org/10.1073/pnas.1715996115

Rowland I, Gibson G, Heinken A et al (2018) Gut microbiota functions: metabolism of nutrients and other food components. Eur J Nutr 57(1):1–24. https://doi.org/10.1007/s00394-017-1445-8

Ruparell A, Inui T, Staunton R, Wallis C, Deusch O, Holcombe LJ (2020) The canine oral microbiome: variation in bacterial populations across different niches. BMC Microbiol 20:42. https://doi.org/10.1186/s12866-020-1704-3

Suchodolski JS (2016) Diagnosis and interpretation of intestinal dysbiosis in dogs and cats. Vet J 215:30–37. https://doi.org/10.1016/j.tvjl.2016.04.011

Summers KL, Arfken AM (2022) The gut Mycobiome and animal health. In: Kogut MH, Zhang G (eds) Gut microbiota, immunity, and health in production animals. The microbiomes of humans, animals, plants, and the environment, vol 4. Springer, Cham. https://doi.org/10.1007/978-3-030-90303-9_6

Thursby E, Juge N (2017) Introduction to the human gut microbiota. Biochem J 474(11): 1823–1836. https://doi.org/10.1042/BCJ20160510

Williams CL, Caraballo-Rodríguez AM, Allaband C, Zarrinpar A, Knight R, Gauglitz JM (2018) Wildlife-microbiome interactions and disease: exploring opportunities for disease mitigation across eco- logical scales. Drug Discov Today Dis Model 28:105–115. https://doi.org/10.1016/j.ddmod.2019.08.012

Xiao L, Estellé J, Kiilerich P, Ramayo Caldas Y, Xia Z et al (2016) A reference gene catalogue of the pig gut microbiome. Nat Microbiol 1:16161. https://doi.org/10.1038/nmicrobiol.2016.161

Xiao Y, Xiang Y, Zhou W, Chen J, Li K, Yang H (2017) Microbial community mapping in intestinal tract of broiler chicken. Poult Sci 96(5):1387–1393. https://doi.org/10.3382/ps/pew372

Chapter 11
Exploring the Impact of Microbial Invasion on the Diseases Involving Gut-Brain Axis in Animals

Pranay Jain, Ram Kumar Pundir, and Ram Prakash Pandey

Abstract The interaction between the host and the animal microbiota has evolved for the host's benefit. Ontogenetically, it is established between the first year of life and throughout the life of the new born baby through the mother. It is horizontally transferred among close family, friends, or neighbours. By metabolizing complex macromolecules that would otherwise be inaccessible nutrients, controlling intestinal motility, neutralizing medicines and carcinogens and enabling visceral awareness; this microecosystem helps to defend the host against a variety of infections. The microbiota-gut-vagus-brain axis, which is observed to function in a bidirectional manner between the brain and the gastrointestinal tract, primarily through the vagus nerve, is essential for maintaining overall body balance and may also play a role in the aetiology of a number of metabolic and mental dysfunctions and disorders in both humans and animals. This chapter explores the understanding of principles of correlatedness between microbiota, animal gut and brain, impact of microbial invasion on diseases involving gut-brain axis in animals and humans and reviews some of the diseases associated with gut-brain axis caused by microorganisms.

Keywords Bidirectional signalling · Central nervous system · Enteric nervous system · Gastrointestinal tract · Microbiota

P. Jain (✉)
Department of Biotechnology, University Institute of Engineering and Technology, Kurukshetra University, Kurukshetra, Haryana, India

R. K. Pundir · R. P. Pandey
Department of Biotechnology, University Institute of Biotechnology, Chandigarh University, Mohali, Punjab, India
e-mail: ramkumar.e11204@cumail.in; ram.e11496@cumail.in

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_11

11.1 Microbiota-Gut-Brain Axis: An Overview

A bidirectional communication has been shown to exist between the microbiota, gut and the brain. This relationship is referred to as the microbiota-brain-gut axis (Dinan and Cryan 2017a, b). The term "microbiota" refers to group of microbes existing in a specific habitat, whereas the meaning of the term "commensals" is colonization of microbes with host without inducing any disease. The relation between the brain and the microbiota of gut and behaviour has the potential to influence new treatment plans for gastrointestinal and neuropsychiatric illnesses brought on by stress (Luczynski et al. 2016).

The intestinal microbiota metabolites influence gastrointestinal function, as they are capable of affecting immune function of gut mucosa, permeability, motility of intestine, sensitivity and enteric nervous system (Forsythe and Kunze 2013). Other multiple mechanisms, i.e. endocrine and neurocranial pathways are associated in gut microbiota brain signalling. The existence of such an axis indicates that, through a dynamic alignment, the central nervous system's vegetative and cognitive functions of the host are affected by the intestinal lumen microbiota that lives there, and that, in turn, brain activity has an impact on the growth, development and reproduction of the microbiota. Bidirectional interactions could be implicated in the pathogenic mechanisms causing the emergence of neurological diseases and could provide procedures for neuroinflammation (Nickel et al. 2003). Figure 11.1 shows interaction between microbiota, gut and brain axis.

The development of metagenomics, in which an unknown bacterium is cultured and identified, increases the information regarding diversity of microbiota in animals and humans. Germ-free mice have the strongest evidence for microbiota role in gut-brain signalling (Lagier et al. 2018). Understanding the bidirectional communication between the microbiota, gut and brain in this regard holds the key to new preventive and therapeutic possibilities as well as major potential effects on global health.

11.1.1 Animal Models for Elucidating Microbiota-Gut-Brain Axis Functioning

Experimental approaches on microbiota-gut-brain axis have so far included animals with pathogenic bacterial infections, the use of animals which are germ-free and animals having exposure to probiotic agents as well as antibiotics. The most convincing evidence for role of these microbes in gut-brain signalling has been found in research on mice raised without any contact to microorganisms, i.e. Germ-free mice. Neurotransmitter, neurotrophic signalling systems, synaptic and neurogenesis are also influenced by the gut microbiota. At various developmental phases, an entire microbiota or a certain group of bacteria can be introduced.

The germ-free model alternatives are necessary and includes treatment using antibiotics in order to decrease the microbial population in animals over the lifespan.

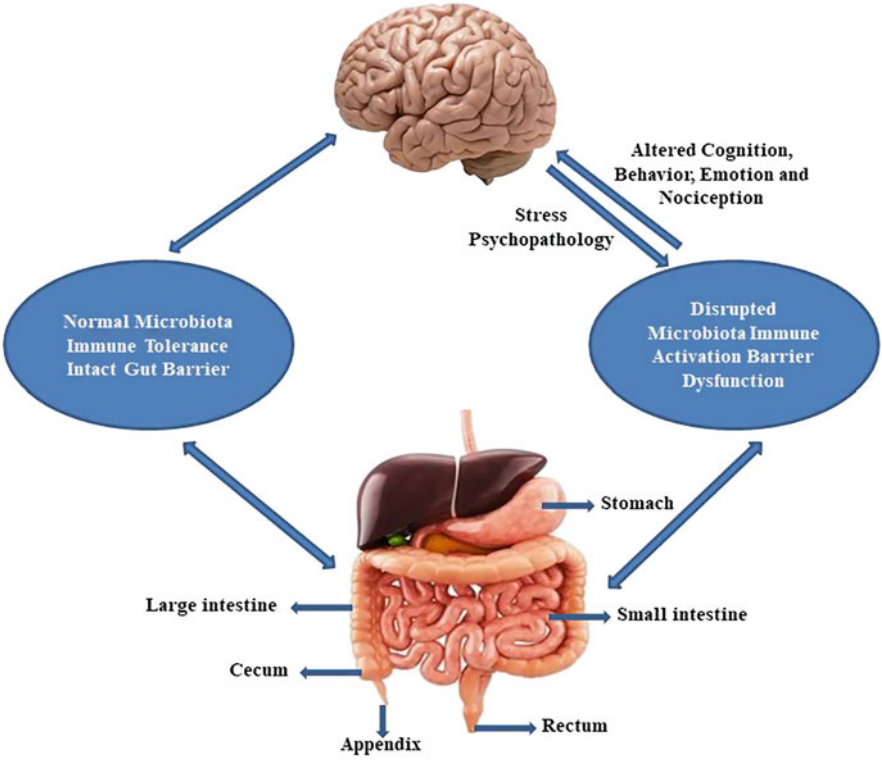

Fig. 11.1 Microbiota-gut-brain axis. Inflammatory mediators and other neuroactive molecules are released into the systemic circulation as a result of an aberrant microbiota, a compromised gut barrier and the activation of the mucosal immune system. These molecules then go to the brain and alter cognition and behaviour. Alternatively, central stressors like stress can impair gut function by affecting the gut microbiota, gut barrier and mucosal immunity

Additionally, the administration of probiotic organisms also increases the level of intestine microflora of the host. Antibiotic administration can alter composition of microbiota in a temporally regulated and clinically relevant manner, making it a potent tool for determining how the gut microbiota affects behaviour (Montiel-Castro et al. 2013). Mayer (2011) has observed that the microbiota-gut-brain axis has a wide range of effects on higher and more complex cognitive functions like motivation and emotion.

11.2 Microbiota-Gut-Brain Axis Components

The gut and the Central Nervous System communication include the neuroendocrine and enteric nervous system, and neuroimmuno signalling pathways (Foster et al. 2017; Cryan et al. 2019). The spinal and vagal sensory nerves' afferent fibres

transmit visceral feedback. They do this by activating polysynaptic entrances to higher order brain areas like the limbic forebrain and hypothalamus (O'Mahony et al. 2011).

11.3 The Blood–Brain Barrier and its Role in Protection from Pathogenic Microorganisms

The brain and gastrointestinal tract emerged from closely related embryonic regions (Dinan and Cryan 2015). Both have in common the blood–brain barrier and the epithelial barrier in intestine. Endothelial cells found in capillaries make up the majority of the blood–brain barrier. Endothelial cells of this barrier are made up of neurons, astrocytes, extracellular matrix and pericytes (Hawkins and Davis 2005). Intracellular tight junction proteins seal the endothelial cells (Braniste et al. 2014). Limiting the water-soluble substances (like transmembrane proteins such as claudins, junction-related molecules as well as cytoplasmic proteins) of paracellular diffusion from blood to the parenchyma of brain ensures proper functioning of neurons (Braniste et al. 2014; Fiorentino et al. 2016). During the neonates' brain development, the blood–brain barrier when intact helps to protect against colonization of microbiota and also shields from microbial metabolites (Knudsen 2004; Braniste et al. 2014).

11.4 The Influence of Gut Microbiota on the Brain

The host gets affected through neural, neuroendocrine, metabolic and immune pathways modulated by the bacteria present in gut including *Bacteroides* and *Firmicutes* (Dinan and Cryan 2017a, b). The key channels of communication between the brain and the microbes in intestine are the tryptophan metabolites, microbial as well as vagal nerve products, like fatty acids having short chain and/or peptidoglycan (Michel and Prat 2016). The brain activities are affected by noradrenergic, modulating serotoninergic, glutamatergic, dopaminergic and GABA (Gamma-aminobutyric acid)-ergic neurotransmission (Fendt et al. 2008; Winter et al. 2018). Microorganisms can influence either metabolism/synthesis of neurotransmitters or can produce these neuroactive substances by themselves. Serotonin is produced by *Candida, Escherichia, Enterococcus* and *Streptococcus* (Dinan et al. 2015); GABA by *Bifidobacterium* and *Lactobacillus* (Reardon 2014). *Lactobacillus* has also been shown to produce acetylcholine; *Bacillus* and *Serratia* produce dopamine; whereas *Escherichia* and *Saccharomyces* are reported to produce norepinephrine (Lyte 2011). Dinan et al. (2015) reported that *Bacillus, Enterococcus, Escherichia, Saccharomyces* or *Streptococcus* increase intestinal noradrenaline. Tryptophan metabolism is controlled by enzymes, which produces serotonin,

kynurenine or indole derivatives (Agus et al. 2018). According to Duerkop et al. (2009), and Forsythe and Bienenstock (2010), the presence of the gut microbiota, as well as pre- and probiotic agents, can have a significant impact on a number of cognitive processes and the levels of circulating cytokines. Additionally, the manipulation of systemic tryptophan, a precursor to the neurotransmitter serotonin and the effect of the vagus nerve's afferent branch are both related to the brain by the gut microbiota (Bercik et al. 2011; Bravo et al. 2011).

11.4.1 The Microbiota-Gut-Brain Axis in Neurobiological Diseases

A fresh perspective on the idea of disease pathophysiology, a paradigm shift in neurobiology is also observed due to the gut microbiota modulating the gut-brain axis (Allen et al. 2017). Pathophysiology of diseases includes psychological stress and inflammation in which microbiota show a very crucial role. De Pama and co-workers (2015) explained about the intestinal microbiota composition and their responsiveness towards stress. Stress is responsible for depression (Willner et al. 2013), schizophrenia (Golofast and Vales 2020), autism spectrum disorder (Theoharides et al. 2019), epilepsy (Kotwas et al. 2017) and migraine (Goadsby et al. 2017); whereas inflammation plays a role in depression (Maes et al. 2016), schizophrenia (Golofast and Vales 2020), autism spectrum disorder (Matta et al. 2019), Parkinson's disease (Rocha et al. 2018), epilepsy (Mazarati et al. 2017) and migraine (Goadsby et al. 2017). Autism spectrum disorder and depression are frequent comorbidities in epilepsy; while depression and migraine frequently co-occur (Mazarati et al. 2017; Amoozegar 2017). Because of the patient's migraines, gastrointestinal conditions such as inflammatory bowel disease are also common (van Hemert et al. 2014).

The relationship between stress and gut microbes is referred to as "leaky gut" phenomenon, which is due to distraction in the gut barrier where the epithelium layer becomes permeable (Foster et al. 2013). Lipopolysaccharide of Gram-negative bacteria is translocated as a result of immune system activation, such as that caused by Toll-like receptors, and the release of pro-inflammatory cytokines, such as IL-6, IFN, CRP and TNF (Maes and Leunis 2008). The memory, emotion and behaviour are predominantly involved in increased levels of pro-inflammatory cytokines and the limbic system is activated by environmental stress (Catani et al. 2013). When the limbic system is activated by the hypothalamic-pituitary-adrenal (HPA) axis, the main stress hormone, cortisol, is released from the adrenal glands and has an impact on a number of human organs, including the brain (Godoy et al. 2018).

11.4.2 Microbiota-Gut-Brain Axis in Depression

The most prevalent mental illness, depression has a complex aetiology that causes multiple anatomical and functional abnormalities in the prefrontal cortex and the hippocampus. In animal models of these conditions, stress is thought to be a trigger for depression (Willner et al. 2013). Anxiety-like behaviour in parallel to depression-like behaviour has been studied in animal models (Neumann et al. 2011). Anxiety and depressive disorders are directly connected with stress, in which the gut microbiota plays a significant role (Foster et al. 2013). The evidence on the gut microbiome's function in regulating changes in behaviour and brain function brought on by stress have also been seen in preclinical investigations.

Clinically, hypothalamic-pituitary-adrenal (HPA) axis dysregulation is linked to depressive episodes, and depressive systems resolve with HPA axis normalization (Heuser et al. 1996; Nickel et al. 2003). The discovery that showed a heightened and adrenocorticotrophin (ACTH) and corticosterone response to restraint stress in germ-free mice when compared to conventionally house-specific pathogen-free mice established a direct link between microbiota and HPA reactivity (Sudo et al. 2004).

11.4.3 Role of Microbiota-Gut-Brain Axis in Autistic Spectrum Disorders

High or low sensitivity, sensory discrimination and motor deficits caused by sensory input are all uncommon behaviours that are part of the autism spectrum. Significant difficulties in social interactions, signs of rigid and repetitive behaviour and communication differences are further characteristics (Bonati et al. 2022). Earlier, in the course of a developmental condition, alterations in intestinal permeability and autistic enterocolitis take place (Theoharides and Doyle 2008). Additionally, biochemical changes that were consistent with the abnormal gut microbiota composition seen in autistic children were discovered by urinary metabolic phenotyping. Recent research suggests that the clinical manifestation of autism may be caused by changes in antigenic load brought on by impaired gut barrier function (de Magistris et al. 2010). Desbonnet et al. (2008) stated that microbiota play an essential role in the regulation of repetitive behaviours as well as the programming and presentation of other regular social behaviours, such as social drive and a desire for social novelty. These behavioural traits are hampered in neurodevelopmental conditions including schizophrenia and autism. According to a study by Kang et al. (2013), autistic children had lower concentrations of the mucolytic bacterium *Akkermansia muciniphila* and *Bifidobacteria* species. Another study found that people with autism had less varied gut microbial makeup, with fewer concentrations of *Prevotella*, *Coprococcus* and unclassified Veillonellaceae (Luna et al. 2017).

11.4.4 Microbiota-Gut-Brain Axis in Alzheimer Disease

The aberrant protein production in and around brain cells causes Alzheimer's. The amyloid protein, which forms plaques surrounding brain cells, is one of the proteins. The second protein, *tau*, creates tangles inside of brain cells (Jiang et al. 2017). According to numerous studies, changes to the gut microbiome may trigger pro-inflammatory cytokines and increase intestinal permeability, which could result in insulin resistance (Bekkering et al. 2013). Lipopolysaccharides (LPSs) and amyloids, which have potent pro-inflammatory and innate-immune actions and activate the system, are released by bacterial species that make up the gut microbiome. Secondary degeneration and decreased peptide cleaning mechanisms follow increased amyloid accumulation, both of which are hallmarks of Alzheimer disease (Zhao et al. 2015). According to Bieschke et al. (2010), diet and particular nutrients may affect the formation or aggregation of amyloid proteins as well as the composition of the gut microbiota.

11.4.5 Microbiota-Gut-Brain Axis in Parkinson Disease

Parkinson's disease mainly affects the motor and the gastrointestinal system (Dinan and Cryan 2017a, b), which further affects the Enteric Nervous System (Lozupone and Knight 2005). Rapid movement of eyes, constipation, sleep behaviour syndrome, hyposmia and depression are observed under Parkinson's disease symptoms (Rodriguez-Violate et al. 2017). One study revealed differential abundances of gut microbial taxa (such as *Akkermansia*, *Anaerotruncus* and *Bacteroides)*, in non-motor symptoms and in Parkinson disease respectively (Heintz-Buschart et al. 2018). The decrease of gut microbes such as *Coprococcus*, *Faecalibacterium*, *Prevotella, Blautia* and *Prevotellaceae* has been observed in Parkinson disease patients by Gerhardt and Mohajeri (2018).

11.4.6 Microbiota-Gut-Brain Axis and Inflammatory Bowel Disease

Two major types of inflammatory bowel illness, ulcerative colitis and Crohn's disease, affect the colon and the small intestine. The mouth, oesophagus, stomach and anus are all impacted by Crohn's disease, in addition to the small and large intestines. The colon and the rectum are most commonly affected in ulcerative colitis (Baumgart and Carding 2007).Microorganisms living in the gastrointestinal tract produce pleiotropic substances that affect the maturation of innate and acquired immunity of host and homeostasis, maintenance of the epithelial barrier function, energy and metabolism, providing defence against pathogenic microbes (Baj et al. 2019). When

compared to conventionally reared controls, it has also been discovered that germ-free animals have significantly different metabolite levels in a variety of biological tissues, including the gut. As a result, they require a higher caloric intake to maintain the same body weight and are more vulnerable to vitamin deficiencies, necessitating dietary supplementation (Sumi et al. 1977). The impact of the bacterial metabolic profile on the homeostasis of the host organism is significantly influenced by changes in the quantities of metabolites with immune-modulatory properties, such as short chain fatty acids, bile acids and tryptophan compounds leading to mucosal inflammation (Lavelle and Sokol 2020). Inadequate protective bacterial metabolite synthesis may negatively affect gut-brain connection, favouring gut-brain diseases related to bowel disease. Allegretti and co-workers (2018) used a novel therapeutic approach with faecal microbiota transplantation in the treatment of neurodegenerative disorders and other diseases. By giving a sample (a faecal specimen from a healthy donor) to the patient through the mouth or the rectum, this approach has produced encouraging and amazingly positive results in patients with recurrent microbial illness. Faecal microbiota transplantation is now a significant medical treatment option.

11.5 Conclusion

The importance of gut microbiota on brain functionality has received increased notice and attention. Numerous studies have demonstrated how stress is influenced by the gut microbiota's composition. Stress reactivity is influenced by the central nervous system and microbiota's reciprocal communication. Numerous studies have demonstrated that the microbiota has an impact on the serotonergic, GABAergic and plasticity-related signalling of CNS systems. The composition of the microbiome can be altered to influence or treat neurological illnesses. Researchers will be able to recognize how the microbiome dysbiosis affects mental disorder when they take into account the gut-brain axis when current and emerging technologies are developed.

References

Agus A, Planchais J, Sokol H (2018) Gut microbiota regulation of tryptophan metabolism in health and disease. Cell Host Microbe 23(6):716–724

Allegretti JR, Kassam Z, Osman M, Budree S, Fischer M, Kelly CR (2018) The 5D framework: a clinical primer for fecal microbiota transplantation to treat *Clostridium difficile* infection. Gastrointest Endosc 87:18–29

Allen AP, Dinan TG, Clarke G, Cryan JF (2017) A psychology of the human braingut-microbiome axis. Soc Personal Psychol Compass 11(4):123

Amoozegar F (2017) Depression comorbidity in migraine. Int Rev Psychiatry 29(5):504–515

Baj A, Moro E, Bistoletti M, Orlandi V, Crema F, Giaroni C (2019) Glutamatergic signaling along the microbiota-gut-brain axis. Int J Mol Sci 20:1482

Baumgart DC, Carding SR (2007) Inflammatory bowel disease: cause and immunobiology. Lancet 369(9573):1627–1640

Bekkering P, Jafri I, van Overveld FJ et al (2013) The intricate association between gut microbiota and development of type 1, type 2 and type 3 diabetes. Expert Rev Clin Immunol 9:1031–1041

Bercik P, Park AJ, Sinclair D, Khoshdel A, Lu J, Huang X, Deng Y, Belnnerhassett PA et al (2011) The anxiolytic effect of *Bifidobacterium longum* NCC3001 involves vagal pathways for gut–brain communication. Neurogastroenterol Motil 23:1132–1139

Bieschke J, Russ J, Friedrich RP et al (2010) EGCG remodels mature a-synuclein and amyloid beta fibrils and reduces cellular toxicity. Proc Natl Acad Sci U S A 107:7710–7715

Bonati, M., Cartabia, M., Clavenna, A. 2022. "Still too much delay in recognition of autism spectrum disorder". Epidemiology and Psychiatric Sciences. Cambridge and New York: Cambridge University Press. 31 (e1): e1

Braniste V, Al-Asmakh M, Kowal C, Anuar F, Abbaspour A, Toth M, Korecka A, Bakocevic N, Kundu LG, Ng P, Gulyas B, Halldin C, Hultenby K, Nilsson H, Hebert H, Volpe BT, Diamond B, Pettersson S (2014) The gut microbiota influences blood-brain barrier permeability in mice. Sci Transl Med 6(263):263

Bravo JA, Forsythe P, Chew MV, Escaravage E, Savignac HM, Dinan TG et al (2011) Ingestion of lactobacillus strain regulates emotional behavior and central GABA receptor expression in a mouse via the vagus nerve. Proc Natl Acad Sci U S A 108:16050–16055

Catani M, Dell'acqua F, Thiebaut de Schotten M (2013) A revised limbic system model for memory, emotion and behaviour. Neurosci Biobehav Rev 37(8):1724–1737

Cryan JF, O'Riordan KJ, Cowan CSM, Sandhu KV, Bastiaanssen TFS, Boehme M, Codagnone MG, Cussotto S, Fulling C, Golubeva AV, Guzzetta KE, Jaggar M, Long-Smith CM, Lyte JM, Martin JA, MolineroPerez A, Moloney G, Morelli E, Morillas E, O'Connor R, Cruz-Pereira JS, Peterson VL, Rea K, Ritz NL, Sherwin E, Spichak S, Teichman EM, van de Wouw M, Ventura-Silva AP, Wallace-Fitzsimons SE, Hyland N, Clarke G, Dinan TG (2019) The microbiota-gut-brain axis. Physiol Rev 99(4):1877–2013

de Magistris L, Familiari V, Pascotto A, Sapone A, Frolli A, Iardino P et al (2010) Alterations of the intestinal barrier in patients with autism spectrum disorders and in their first-degree relatives. J Pediatr Gastroenterol Nutr 51(4):418–424

Desbonnet L, Garrett L, Clarke G, Bienenstock J, Dinan TG (2008) The probiotic *Bifidobacteria infantis*: an assessment of potential antidepressant properties in the rat. J Psychiatr Res 43:164–174

Dinan TG, Cryan JF (2015) The impact of gut microbiota on brain and behaviour: implications for psychiatry. Curr Opin Clin Nutr Metab Care 18(6):552–558

Dinan TG, Cryan JF (2017a) The microbiome-gut-brain axis in health and disease. Gastroenterol Clin N Am 46(1):77–89

Dinan TG, Stilling RM, Stanton C, Cryan JF (2015) Collective unconscious: how gut microbes shape human behaviour. J Psychiatr Res 63:1–9

Dinan TG, Cryan JF (2017b) Gut-brain axis in 2016: brain-gut-microbiota axis—mood, metabolism and behaviour. Nat Rev Gastroenterol Hepatol 14(2):69

Duerkop BA, Vaishnava S, Hooper LV (2009) Immune responses to the microbiota at the intestinal mucosal surface. *Immunity* 31:368–376. https://doi.org/10.1016/j.immuni.2009.08.009

Fendt M, Schmid S, Thakker DR, Jacobson LH, Yamamoto R, Mitsukawa K, Maier R, Natt F, Hüsken D, Kelly PH, McAllister KH, Hoyer D, van der Putten H, Cryan JF, Flor PJ (2008) mGluR7 facilitates extinction of aversive memories and controls amygdala plasticity. Mol Psychiatry 13(10):970–979

Fiorentino M, Sapone A, Senger S, Camhi SS, Kadzielski SM, Buie TM, Kelly DL, Cascella N, Fasano A (2016) Blood-brain barrier and intestinal epithelial barrier alterations in autism spectrum disorders. Mol Autism 7:49

Forsythe P, Bienenstock J (2010) Immunomodulation by commensal and probiotic bacteria. Immunol Investig 39:429–448. https://doi.org/10.3109/08820131003667978

Forsythe P, Kunze WA (2013) Voices from within: gut microbes and the CNS. Cell Mol Life Sci 70(1):55–69

Foster JA, McVey KA, Neufeld A (2013) Gut-brain axis: how the microbiome influences anxiety and depression. Trends Neurosci 36(5):305–312

Foster JA, Rinaman L, Cryan JF (2017) Stress & the gut-brain axis: regulation by the microbiome. Neurobiology of Stress 7:124–136

Gerhardt S, Mohajeri M (2018) Changes of colonic bacterial composition in Parkinson's disease and other neurodegenerative diseases. Nutrients 10(6):708

Goadsby PJ, Holland PR, Martins-Oliveira M, Hoffmann J, Schankin C, Akerman S (2017) Pathophysiology of migraine: a disorder of sensory processing. Physiol Rev 97(2):553–622

Godoy LD, Rossignoli MT, Delfino-Pereira P, Garcia-Cairasco N, de Lima Umeoka EH (2018) A comprehensive overview on stress neurobiology: basic concepts and clinical implications. *Frontiers in Behavioural*. Neuroscience 12:127

Golofast B, Vales K (2020) The connection between microbiome and schizophrenia. Neurosci Biobehav Rev 108:712–731

Hawkins BT, Davis TP (2005) The blood-brain barrier/neurovascular unit in health and disease. Pharmacol Rev 57(2):173–185

Heintz-Buschart A, Pandey U, Wicke T, Sixel-Doring F, Janzen A, Sittig-Wiegand E et al (2018) The nasal and gut microbiome in Parkinson's disease and idiopathic rapid eye movement sleep behaviour disorder. Mov Disord 33(1):88–98

Heuser IJ et al (1996) Pituitary-adrenal-system regulation and psychopathology during amitriptyline treatment in elderly depressed patients and normal comparison subjects. Am J Psychiatr 153:93–99

Jiang C, Li G, Huang P, Liu Z, Zhao B (2017) The gut microbiota and Alzheimer's disease. J Alzheimers Dis 58(1):1–15

Kang DW, Park JG, Ilhan ZE, Wallstrom G, LaBaer J, Adams JB et al (2013) Reduced incidence of Prevotella and other fermenters in intestinal microflora of autistic children. PLoSOne 8(7): e68322

Knudsen EI (2004) Sensitive periods in the development of the brain and behavior. J Cogn Neurosci 16(8):1412–1425

Kotwas I, McGonigal A, Bastien-Toniazzo M, Bartolomei F, MicoulaudFranchi JA (2017) Stress regulation in drug-resistant epilepsy. Epilepsy Behav 71(Pt A):39–50

Lagier JC, Dubourg G, Million M, Cadoret F, Bilen M, Fenollar F, Levasseur A, Rolain JM, Fournier PE, Raoult D (2018) Culturing the human microbiota and culturomics. Nat Rev Microbiol 16:540–550

Lavelle A, Sokol H (2020) Gut microbiota-derived metabolites as key actors in inflammatory bowel disease. Nat Rev Gastroenterol Hepatol 17:223–237

Lozupone C, Knight R (2005) UniFrac: A new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 71(12):8228–8235

Luczynski P, McVey Neufeld K-A, Oriach CS, Clarke G, Dinan TG, Cryan JF (2016) Growing up in a bubble: using germ-free animals to assess the influence of gut microbiota on brain and behaviour. Int J Neuropsychopharmacol 19(8):1–17

Luna RA, Oezguen N, Balderas M, Venkatachalam A, Runge JK, Versalovic J et al (2017) Distinct microbiome-neuroimmune signatures correlate with functional abdominal pain in children with autism spectrum disorder. Cell Mol Gastroenterol Hepatol 3(2):218–230

Lyte M (2011) Probiotics function mechanistically as delivery vehicles for neuroactive compounds: Microbial endocrinology in the design and use of probiotics. Bioessays **33**(**8**):574–581

Maes M, Leunis JC (2008) Normalization of leaky gut in chronic fatigue syndrome (CFS) is accompanied by a clinical improvement: effects of age, duration of illness and the translocation of LPS from gram-negative bacteria. Neuroendocrinol Lett 29(6):902–910

Maes M, Nowak G, Caso JR, Leza JC, Song C, Kubera M, Klein H, Galecki P, Noto C, Glaab E, Balling R, Berk M (2016) Toward omics-based, systems biomedicine, and path and drug

discovery methodologies for depression-inflammation research. Mol Neurobiol 53(5): 2927–2935
Matta SM, Hill-Yardin EL, Crack PJ (2019) The influence of neuroinflammation in autism spectrum disorder. Brain Behav Immun 79:75–90
Mayer EA (2011) Gut feelings: the emerging biology of gut–brain communication. Nat Rev Neurosci 12:453–466. https://doi.org/10.1038/nrn3071
Mazarati AM, Lewis ML, Pittman QJ (2017) Neurobehavioral comorbidities of epilepsy: role of inflammation. *Epilepsia* 58(3):48–56
Michel L, Prat A (2016) One more role for the gut: microbiota and blood brain barrier. Ann Transl Med 4(1):15
Montiel-Castro AJ, González-Cervantes RM, Bravo-Ruiseco G, Pacheco-López G (2013) The microbiota-gut-brain-axis: neurobehavioural correlates, health and sociality. Front Integr Neurosci 7:70. https://doi.org/10.3389/fnint.2013.00070
Neumann ID, Wegener G, Homberg JR, Cohen H, Slattery DA, Zohar J, Olivier JD, Mathe AA (2011) Animal models of depression and anxiety: what do they tell us about human condition? Prog Neuro-Psychopharmacol Biol Psychiatry 35(6):1357–1375
Nickel T et al (2003) Clinical and neurobiological effects of tianeptine and paroxetine in major depression. J Clin Psychopharmacol 23:155–168
O'Mahony SM, Hyland NP, Dinan TG, Cryan JF (2011) Maternal separation as a model of brain-gut axis dysfunction. Psychopharmacology 214(1):71–88
Reardon S (2014) Gut-brain link grabs neuroscientists. Nature 515(7526):175–177
Rocha EM, De Miranda B, Sanders LH (2018) Alpha-synuclein: pathology, mitochondrial dysfunction and neuroinflammation in Parkinson's disease. Neurobiol Dis 109(Pt B):249–257
Rodriguez-Violate M, Zeron-Martinez R, Cervantes-Arriaga A, Corona T (2017) Who can diagnose Parkinson's disease first? Role of pre-motor symptoms. Arch Med Res 48(3):221–227
Sudo N et al (2004) Post-natal microbial colonization programs the hypothalamic-pituitary-adrenal system for stress response in mice. J Physiol 558:263–275
Sumi Y, Miyakawa M, Kanzaki M, Kotake Y (1977) Vitamin B-6 deficiency in germ-free rats. J Nutr 107:1707–1714
Theoharides TC, Kavalioti M, Tsilioni I (2019) Mast cells, stress, fear and autism spectrum disorder. Int J Mol Sci 20(15):3611
Theoharides TC, Doyle R (2008) Autism, gut-blood-brain barrier and mast cells. J Clin Psychopharmacol 28(5):479–483
van Hemert S, Breedveld AC, Rovers JM, Vermeiden JP, Witteman BJ, Smits MG, de Roos NM (2014) Migraine associated with gastrointestinal disorders: review of the literature and clinical implications. Front Neurol 5:241
Willner P, Scheel-Krüger J, Belzung C (2013) The neurobiology of depression and antidepressant action. Neurosci Biobehav Rev 37(10 Pt 1):2331–2371
Winter G, Hart RA, Charlesworth RPG, Sharpley CF (2018) Gut microbiome and depression: what we know and what we need to know. Rev Neurosci 29(6):629–643
Zhao Y, Dua P, Lukiw WJ (2015) Microbial sources of amyloid and relevance to amyloidogenesis and Alzheimer's disease (AD). J Alzheimer's Dis Parkinsonism 5:177

Chapter 12
Genome-Based Prediction of Bacterial Antibiotic Resistance

Deepali Kalambhe, Lokesh K. M., Gourab Basak, Abhilash Jadhao, and Sumeet Singh

Abstract Genome-based techniques, especially whole-genome sequencing (WGS), present a huge potential to predict antimicrobial resistance in the microbes. The advancement in the inexpensive DNA sequencing technology, bioinformatics tools and handy online databases on nucleotide sequence has transformed the entire diagnostic microbiology and bacterial investigation. Genome sequencing in conjugation with the online bioinformatics tools helps in predicting real-time AMR determinants. This approach allows establishing global pathogen surveillance and AMR tracking based on genomics which is essential to combat, control and prevent the increasing threat of AMR. Tools genome-based surveillance tools are either available at public genome data domains or can be operated locally. Public database centres such as NCBI and European Nucleotide Archive (ENA) allow online submission of nucleotide sequencing data along with phenotypic antimicrobial susceptibility data. However, there is a need for optimization of databanks as well as phenotypic predictions based on the genomic data. This chapter discusses the latest genome-based techniques, bioinformatics tools and genomic databases for predicting antimicrobial resistance (AMR).

Keywords Whole-genome sequencing · Bioinformatics · Antimicrobial resistance · Next-generation sequencing · In silico analysis

D. Kalambhe (✉) · L. K. M. · G. Basak
Centre for One Health, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

A. Jadhao
Department of Veterinary Pathology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

S. Singh
Department of Veterinary Microbiology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_12

12.1 Introduction

With the discovery of penicillin as the world's first antibiotic in the late 1920s, it has made a renaissance in the medical therapeutics and science; subsequently discovery of various antibiotics from time to time had saved millions of lives across the globe. But alike every coin has two opposite sides, antibiotic resistance also came up along with it in the path within a couple of years. Antimicrobial resistance (AMR) is an emerging and rapidly growing public health concern. According to Food and Agriculture Organization of the United States, AMR is the microbe's ability to persist and continuously grow in presence of the antimicrobial compounds that are intended to inhibit or to kill them (https://www.fao.org/antimicrobial-resistance/background/what-is-it/en/). The transmission of AMR strains of microbes (bacteria) from animals to humans is well documented which are termed as zoonotic AMR strains (Economou and Gousia 2015). Thus, AMR threatens the practice of medicine inferring animal and/or human health risks. Besides, it also possesses implications in food safety and security. Reports highlight that the AMR organisms in United States are responsible for more than two million infective cases along with 23,000 mortalities per annum (CDC 2013). Similarly, Europe is no behind with 25,000 live-loses annually keeping in pace with the former nation (Gelband et al. 2015). Globally, 0.7 million annual deaths occur presently which is expected to touch ten million by 2050 ceasing 66 trillion financial wealth. Moreover, it is been three long decades since the introduction of any new antibiotic.

12.2 Why Antimicrobials Are Used in Animal Production and How AMR Is Knocking?

Diseases are the main reasons for the use of antimicrobials. Though if modified environmental hygiene, proper balanced nutrition and more importantly husbandry and management practices are espoused unanimously, the animals can be preven

from any infectious or metabolic diseases in turn exempting them from any antimicrobials. Besides, the major concern these days is about the use of such chemicals as growth promoters and production enhancers, which gradually keep on bioaccumulation producing antimicrobial resistance (Fig. 12.1). In this regard, limited access to health professionals, oversight and regulation of their use and incomplete completion of drug regiment amplify the condition many folds whereas restricted training provisions for these experts time to time add a cheery on the top. As a result, owners fill the gap with over-the-counter drugs, which really pose a risk impose in the health of animals, man as well as in the environment. These may sometimes also because of substandard and/or falsified drugs which fail to fulfil its target in place encourage the microbes to get acclimatize with the condition provided. Besides, lack of knowledge, awareness and proper use help in increasing height of the graph.

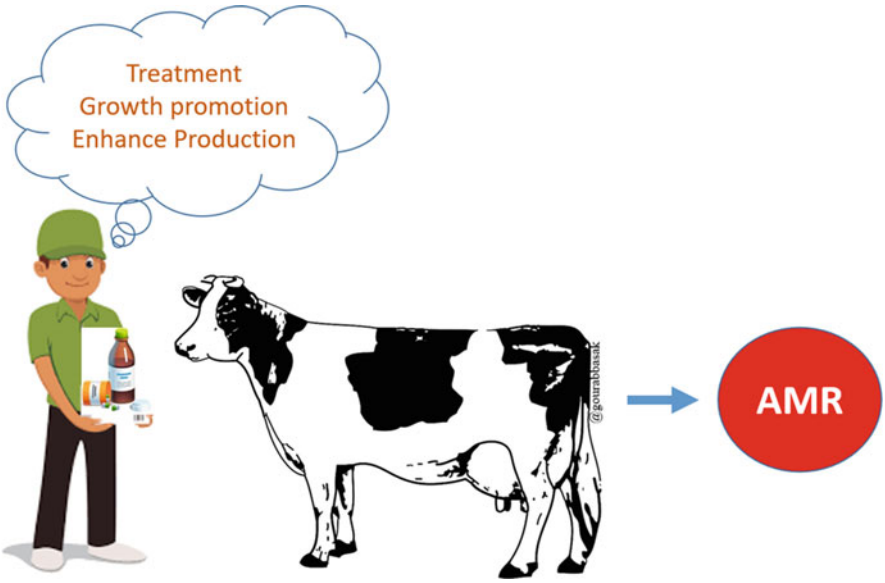

Fig. 12.1 Schematic representation of the drivers behind antimicrobials usage in animal production and their impact

As a result, detection of AMR organisms and the responsible genes became mandatory to arrest the proficiency of AMR contractions. Monitoring of antimicrobial resistance in foodborne pathogens isolated from clinical, food and environmental samples again becomes very important. As because monitoring aid in recognizing and mitigating resistant strains spread from animals to humans inferring public health risk. As the morbidity and economic burden are increasing with the rise in resistance rates, therefore, to guide treatment decisions, precise detection of antibiotic resistance is required. Currently, approaches including phenotypic detection and rapid genomic detection methods like PCR for resistance determinants are in the limelight. Clinical laboratories usually use culture-based antimicrobial susceptibility testing (AST) as their principal approach. But, only or solely phenotypic detection would not help in the long run as more precision can be drawn efficiently through the genotypic methods. In this context, the genome-based detection of Antimicrobial Resistance/Susceptibility Testing offers the potential benefit for rapid, reliable and precise predictions of every known resistance phenotype for a strain.

12.3 Whole-Genome Sequencing for Antimicrobial Susceptibility Testing (WGS-AST)

Whole-genome sequencing for antimicrobial susceptibility testing enabled to assess genes accounting AMR, their location along with their potentiality for multidrug resistance and rapid dissemination (Karp et al. 2017). Single-nucleotide variants,

insertions/deletions, copy number alterations and significant structural variants can easily be detected by whole-genome sequencing. In silico examination for the presence of antimicrobial genes can be done using software and databases including Resistance Gene Identifier in Comprehensive Antimicrobial Resistance Database (CARD), Antibiotic Resistance Database and Resistance Gene Finder (ResFinder). CDC, FDA, FSIS and ARS actively collaborate on nationwide AMR surveillance for near 20 years in the National Antimicrobial Resistance Monitoring System in the United States (Karp et al. 2017).

Previously used Sanger sequencing was highly accurate for relatively shorter DNA fragments, for the longer DNA stretches, the process was time-consuming and involved multiple reactions (Sanger et al. 1977). It took several years to sequence the bacterial genome and employed millions of dollars. With the advent of next-generation sequencing in the early 2000, the revolutionary approach linked DNA sequencing to food safety and public health surveillance on a routine basis. Whole-genome sequencing is an advanced and comprehensive method for determining the complete DNA sequence of an organism. Using next-generation sequencing (NGS) techniques, like Illumina and Nanopore, the sequence of complete chromosomal, plasmid and mitochondrial DNA can be determined in a single reaction and much less time. The entire bacterial genome can be sequenced in small random fragments (1000 bp) multiple times in a single reaction (Vincent et al. 2013). The complete DNA sequence is determined with the use of state-of-the-art bioinformatics tools. WGS provides a very high-resolution base to base view of the genome. Continuous advancement in biotechnology, bioinformatics and information technology enhances capability of using NGS to augment food safety and public health (Allard et al. 2019). This technique can determine a large amount of data in a short time on a routine basis and facilitates to maintain a database for further reference.

- *Base to base (Single-Nucleotide Polymorphism, SNPs)*
- *Gene to gene (Multilocus Sequence Typing, MLST)*

Base to base comparison of the test strain with the reference strain provides the nucleotide difference at specific positions owing to the genetic mutations. SNPs occur throughout the genome in both coding and non-coding regions. The SNP identification reference strain choice is quite significant and can be customized to any given situation (i.e. closely related to the outbreak strain) providing a more precise SNP difference-assessment in an outbreak setting. SNPs profiles of all the isolates are compared pairwise and displayed in the phylogenetic tree. This approach uses all the information of the genome (coding and non-coding regions) and provides greater accuracy for the reconstruction of a strain phylogeny (Pettengill et al. 2014).

Gene to gene approach works by assessing the sequence variation in the coding region of the genes. The accessibility of number and nature of the genes to be assessed to any given situation makes this approach more flexible. In this, genes identified in test strain are compared against the reference databases of genes with all known gene variants of the species. Each unique allele sequence is given a number and the genome is compared based on allele numbers. This approach cannot assess

the variation in non-coding regions. It is more popular in clinical settings as fewer bioinformatics skills are required in this analysis approach.

SNPs provide precise information and the flexibility to choose more closely associated reference strains for precise assessment of relatedness. It is more discriminatory as it assesses non-coding regions also. But for practical purposes, both the approaches are equally discriminatory and epidemiologically concordant (Brown et al. 2018).

Various DNA genome databases are available to share information regarding outbreaks identification and diagnostics of causative agents, which facilitates the early recognition and investigation of international foodborne outbreaks. PulseNet by Centre for Disease Control and Prevention, Food Safety Inspection System (FSIS) by United States Department of Agriculture and GenomeTrakr by U.S. Food and Drug Administration are the networks to utilize whole-genome sequencing for pathogen identification. These databases collect and share genomic and geographic data from foodborne pathogens which can be accessed by researchers and public health officials for real-time comparison and analysis of an outbreak (Stevens et al. 2017). These networks promise speed for foodborne illness outbreak investigations and reduce foodborne illnesses and deaths. Data are stored in the standardized method so that the large volume is greatly reduced to exchange with any part of the world, and least post-processing is required ensuring fast comparison of data from databases in different regions of the world.

WGS has been used in routine public health surveillance since 2014 for *Listeria monocytogenes*, surveillance for *Campylobacter* outbreaks incorporated in 2018, followed by Shiga-toxin producing *E. coli* and *Salmonella* in 2019. PublicNet, FSIS, GenomeTrakr work along with other public health partners to make improvements in hazard detection characterization methods. These networks use WGS for isolation, characterization and surveillance of outbreaks to detect and prevent contamination events and follow foodborne illness outbreaks (Brown et al. 2018).

Genome sequence possesses many advantages as well as has the potential symbiotic interaction between genomics and phenotypic-based AST. Following selective cultivation of the bacterium of interest from a clinical sample, WGS-AST is performed. It is also possible to perform AST after direct shotgun sequencing of clinical samples (Meta genomic-AST). Due to the presence of potentially low amount of pathogen of interest relative to the host DNA, metagenomics is more difficult, expensive and prone to false-negative results. As DNA sequencing is easier and faster than acquiring enough culture growth for phenotypic assessment, slow-growing or difficult-to-culture bacteria (such as *Mycobacterium tuberculosis*) are the key early targets for metagenomic-AST (Doyle et al. 2018; Votintseva et al. 2017). WGS-AST can determine the antibiotic resistance phenotypes of the entire genome simultaneously and phenotypes where multiple loci contribute can be easily screened, unlike culture-based AST or nucleic acid amplification tests (NAATs). The later are often limited by the number of resistant phenotypes that can be determined from one test (except for multiplex PCRs). The genome sequence data are digitally saved and can be queried for additional purposes once it is obtained (Feijao et al. 2018) (Fig. 12.2).

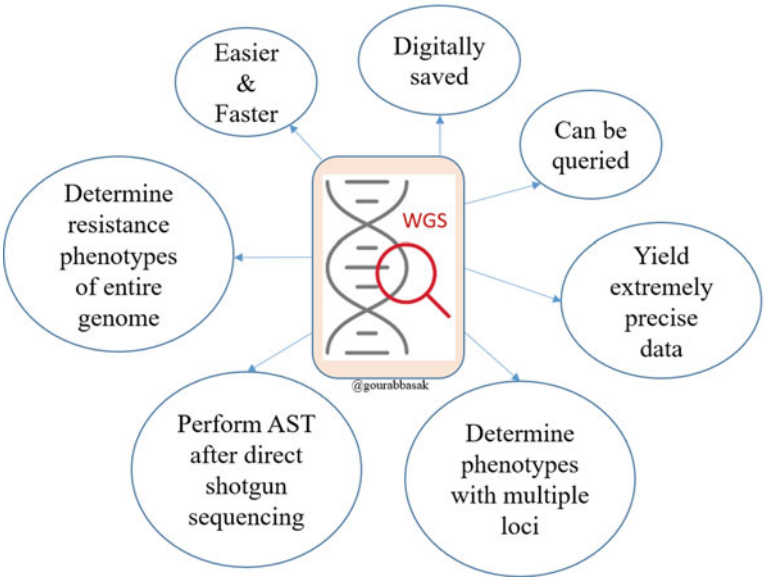

Fig. 12.2 Whole-genome sequencing as tool for antimicrobial susceptibility testing

Genomes can be sequenced to extremely high depths, yielding extremely precise sequence data. Unlike NAATs, template amplification does not rely on primer specificity, lowering the risk of false-negative results. The collection of genomes in clinical laboratories has resulted in a data source that may be utilized to track disease evolution (Gardy and Loman 2018). If new antibiotic resistance loci are discovered, these databases may be searched right away to see how long these genes have been circulating and how they got into the clinical use.

12.4 Next-Generation Sequencing (NGS) Technologies Driving WGS-AST

Next-generation sequencing (NGS) is a high-throughput, low-cost and quick second-generation sequencing technology while whole-genome sequencing (WGS) is a comprehensive method of analyzing the entire genomic DNA of a cell at a single time by using sequencing techniques such as Sanger sequencing, shotgun approach or high-throughput NGS sequencing. Second-generation devices, such as the Illumina sequencing-by-synthesis technology, drastically lowered the cost of data generation allowing for large-scale sequencing of thousands of pathogen

genomes and the application of shotgun metagenomics for clinical diagnosis. Illumina sequencing reads are short (300 bp), paired-end and have a low per-base error rate (usually 0.1%). De novo assembly generally results in genomes fragmented into many contigs and collapsed repeat regions, despite the fact that Illumina sequencing provides for extensive shotgun coverage with high consensus accuracy.

Longer reads are produced using third-generation single-molecule sequencing, as demonstrated by Oxford Nanopore Technology (ONT) and Pacific Biosciences (PacBio) technologies. However, extensive read genome assemblies include fewer gaps and generally span long repeat regions, allowing complicated structural features like tandem repeats and nested insertions to be resolved (Giordano et al. 2017). Third-generation technologies have a greater cost per base and higher per-base error rates (5 to 15%) than Illumina, despite improvements in chemistry and base calling algorithms narrowing the gap (Rhoads and Au 2015; Lu et al. 2016). As existing technologies mature and become more cost-effective, and new approaches emerge, the future of clinical sequencing is continuously shifting. Illumina is currently the most popular WGS-AST platform. However, the estimated minimum cost of $80 per genome is still too high for clinical laboratories to use on a regular basis.

12.5 WGS-AST Based on Searching Catalogues of Resistance Loci

Using a "rules-based" classification based on the presence of one or more known antimicrobial resistance (AMR) genes or mutations is the easiest technique. Cross-referencing the genomic sequence against databases of antibiotic resistance determinants is required for this. The majority of the databases were created via curation of the literature on molecular genetic studies that link antibiotic resistance phenotypes to genes (Xavier et al. 2016).

12.5.1 Multispecies Database

CARD	McArthur et al. (2013), Jia et al. (2017)
Resfinder	Zankari et al. (2012)
Pointfinder	Zankari et al. (2017)
ARG-ANNOT	Gupta et al. (2014)
MEGARes	Liu and Pop (2009)
Resfams	Lakin et al. (2017)
RAST	Gibson et al. (2015)
Bacterial antimicrobial resistance reference gene database	BARRGD, https://www.ncbi.nlm.nih.gov/bioproject/313047

There are also databases created specifically for single species, such as Dream TB (Sandgren et al. 2009) and MUBII-TB-DB (Flandrois et al. 2014) for Mycobacterium tuberculosis. The data produced at two phases in the next-generation sequencing method, raw sequence data and assembled contigs, are used by software tools for rules-based antibiotic resistance catalogue matching. In terms of speed and precision, each has its own set of trade-offs. There is good agreement between what is known about the genetic basis of resistance and the resistance phenotype for many species and antibiotic resistance phenotypes. For numerous characteristics across several pathogen species, rules-based WGS-AST has been proven to have high sensitivity and specificity (95 per cent) (Bradley et al. 2015; Clausen et al. 2016; Mason et al. 2018). Despite the fact that the number of strains examined and the within-species genetic diversity of the test set varied greatly between investigations (Kos et al. 2015). Miotto et al. (2017) divided the predictive power of M. tuberculosis mutations into high, moderate and minimal confidence in an exhaustive investigation of the genetic basis of resistance in M. tuberculosis. As a result, rule-based approaches may not always be enough for accurate WGS-AST.

12.6 Advantages and Disadvantages of Whole-Genome Sequencing

SN	Advantages	Disadvantages	Authors
1	Provides huge genomic data in a single assay	High cost and resources	Ellington et al. (2017), Meienberg et al. (2016)
2	AMR bacteria can be typed and tracked using unique allele profiles	Processing and storing a large amount of data	Hendriksen et al. (2019), Meienberg et al. (2016), Gonzaga-Jauregui et al. (2012)
3	Help to investigate a drug-resistant foodborne and inconsistent resistance patterns among indistinguishable PFGE types of *Salmonella* serovar	Sanger sequencing is required to validate genetic variations	Gong et al. (2018), Gonzaga-Jauregui et al. (2012)
4	Reveal the co-carriage of individual genes generating diverse MDR patterns, allelic trends over time, horizontal transfer and distribution	Large number of variants can be detected in non-coding areas, which may or may not be relevant	Ellington et al. (2017), Gong et al. (2018), Gonzaga-Jauregui et al. (2012)
5	Defines MDR as resistance to three or more drug classes	Physicians are not familiar in interpreting genomic data	Ng and Kirkness (2010)
6	Sequence-based surveillance allows more precise definition of multidrug resistance (MDR) when compared phenotypic method	The function of the majority of gene in the human genome is unclear hence, much of the "knowledge" found in a human genome sequence is currently useless	Hendriksen et al. (2019), Ng and Kirkness (2010)

(continued)

SN	Advantages	Disadvantages	Authors
7	Sequencing allows identification of a single-nucleotide variants, insertions/deletions, copy number alterations and significant structural variants and monogenic disease	Enormous amount of data are generated. Policies and security procedures to protect the privacy and security of this data are still being developed	Abdelbary et al. (2017), Ng and Kirkness (2010)
8	Used to diagnose genetic mutations and to identify genetic carriers of recessive disorders like cystic fibrosis	Requires high informatics capacity and special software	Katsanis and Katsanis (2013), Ng and Kirkness (2010)
9	WGS in cancer research allow the identification of genetic drivers of tumours and new biological therapies	Large incidental findings may increase the risk of overdiagnosis	Zhao et al. (2019), Mazzarotto et al. (2020)
10	Both coding and non-coding variations are detected	Potential inclusion of non-validated genes in genetic testing	Meienberg et al. (2016), Mazzarotto et al. (2020)
11	Detects structural variants	False-positive findings	Gong et al. (2018), Gonzaga-Jauregui et al. (2012), Mazzarotto et al. (2020)

WGS in cancer research could lead to the discovery of new biological therapies and genetic causes of tumours.

12.7 Antibiotic Resistance Techniques Based on Bioinformatics

Bioinformatics uses computer software tool such as sequence and structural alignment application that develop extrapolations and process sequence data as reads or assemblies for finding novel biology from plethora of biological data generated from gene sequences, cell populations, or protein samples (Luscombe et al. 2001). Bioinformatics has become an essential approach for the consolidation of knowledge on antibiotic resistance. Bioinformatics approaches such as molecular docking are commonly used to evaluate ligand–protein interactions and to quantify binding energy during the docking process. It can be utilized to investigate the links between traditional mathematical modelling and omic scope predictions, as well as specific features of the immune system (Rapin et al. 2010). Swiss-model (online homology modelling), Autodock Vina (ligand-protein docking), Avogadro (ligand energy minimization) and Chimera (3D docked complexes) are a few examples of regularly used bioinformatics tools for understanding the antimicrobial profile of bacteria. Bioinformatics method for reducing antibiotic resistance has expanded in recent years, and it now includes bioinformatics techniques based on whole-genome sequencing (Ndagi et al. 2020).

Now a days the whole-genome sequencing (WGS) of pathogens is in vogue due to its easy accessibility, rapid increase in output, rapid analysis and reduced cost (Quainoo et al. 2017). With the advancement in sequencing technologies and analysis tools, the genome sequencing offers a suitable framework for scientific advancement, notably in biomolecular modelling and medication creation, with a focus on antibiotic resistance (Gwinn et al. 2019). Genotyping technologies enable better understanding of disease transmission hence helpful for epidemic management (Quainoo et al. 2017). The whole-genome analysis of bacteria by sequencing provides better insight of related lineages and outbreaks in hospitals (Biswas et al. 2008; Quainoo et al. 2017).

Deoxyribonucleic acid (DNA) sequencing is an excellent platform for protein modelling and drug development (Blundell et al. 2006). Advancements in genome sequencing, protein expression, high-throughput crystallography and nuclear magnetic resonance (NMR) have revolutionized the possibilities for using protein three-dimensional structures to speed drug development. Structural biology and bioinformatics have well-established functions in target identification of the remarkable bacteria resistance in our environment (Blundell et al. 2006). High-throughput structure determination tools offer effective strategies for combating bacterial resistance as it permits screening of complicated bacterium proteins for the identification of lineages, sequence alignment and three-dimensional modelling structures.

12.8 In Silico Analysis of Serovar, Serogroup and Antigenic Profile

With the progression of genetic studies, the inclination of microbiologists have increased for the whole-genome sequencing for the purpose of genotyping just like the molecular serotyping which has replaced the traditional serotyping method. Presently, the multilocus sequence typing (MLST) and serovar-specific gene markers or DNA fragments are more popular in silico serovar prediction techniques for identifying the genes expressing surface antigen. However, these serovar-specific gene markers or DNA fragments can distinguish only a few serovars. This shortcoming was overcome by Zhang et al. (2019) who developed in silico serovar prediction technique that compares 1089 genomes covering 106 serovars to a set of 131 serovar-specific gene markers. According to available literature, the method best suits as a good diagnostic tool for culture-independent and metagenomics methodologies, and also as an alternative for confirming other genome-based investigations. This set of bioinformatics procedures is beneficial for identifying a certain type of gene marker and may help in the development of more cost-effective molecular assays for detecting specific gene markers of all major serovars.

12.9 In Silico Plasmid Identification

Plasmids are the mobile genetic element that are placed external to chromosomal DNA and are capable of autonomous replication. Plasmids are either circular or linear in nature and are abundantly found in microorganisms especially in the bacterial and archaeal domains (Jesus et al. 2019). Plasmids often carry genes for virulence and resistance and by virtue of its mobility it serves as "vehicle" for the transport of genetic information between the bacterial species and genus (Frost et al. 2005). Thus, plasmids are important for the acquisition of virulence traits and spread of antibiotic resistance (Jesus et al. 2019). The growing incidence of plasmid-mediated microbial resistance against the commonly used therapeutic drugs is a big concern for the spread of resistance across the human and veterinary healthcare settings. Therefore, it is absolutely necessary to investigate the molecular epidemiology of plasmids along with the molecular epidemiology of different bacterial strains. There are few online available tools such as cBar, PLACNET, plasmidSPAdes and Recycler (Zhou and Xu 2010; Lanza et al. 2018; Antipov et al. 2016; Rozov et al. 2017) that may be used to extract and assemble plasmids data for specific markers or unique characteristic sequences from high-throughput sequencing (HTS) data. Some other software tools such as plasmidFinder and MOB-suite Plasmid Profiler are also available to reconstruct or detect plasmids in HTS data; however, these are quite difficult for the users to understand the list of hits and evaluate the impact of these alleged plasmids on the host bacteria (Carattoli et al. 2017; Robertson and Nash 2018; Zetner et al. 2017).

The National Centre for Biotechnology Information (NCBI) has approximately 13,924 reference plasmid sequence (RefSeq) entries stored in its data bank though, with a dearth of essential tools for retrieving these massive amounts of plasmid sequence data (Jesus et al. 2019; O'Leary et al. 2016). Plasmid Atlas (pAT-LAS) 102, on the other hand provides an easily accessible visual analytics tool for users to explore the NCBI database for RefSeq plasmids for plasmid identification from HTS data (Jesus et al. 2019). The de novo annotations-based CARD, ResFinder, Virulence Factors Database (VFDB) and PlasmidFinder, pATLAS allows users to envisage and reconnoitre the metadata associated with all plasmids available in NCBI's RefSeq database, as well as their putative antibiotic resistance and virulence genes and plasmid families (Wang et al. 2015; Zankari et al. 2012; Chen et al. 2005; Carattoli et al. 2017 and Jesus et al. 2019).

12.10 Metagenomics for Antimicrobial Surveillance

Surveillance of antimicrobial resistance (AMR) mainly relies on the passive reporting of laboratory generated data on phenotypes of microorganisms. The Danish Monitoring System (DANMAP) (https://www.danmap.org.) is one such surveillance type that provides data on antimicrobial resistance gene pattern based

on the molecular studies generated from the laboratory. However, this type of antimicrobial gene surveillance does not cover all the relevant information as it is confined to a selected spectrum of microorganism. Microbial culture-based techniques can provide a good insight to antimicrobial resistance organism by allowing the whole-genome sequencing of resistance strains of organisms. However, these techniques are tedious and the applicability is limited to only a few genes of interest from the easily culturable organisms. Culture-based methods are not useful to study the antimicrobial resistance profile of unculturable microorganisms. The metagenomics overcome this hurdle as it targets the sequence analysis of genomic material directly extracted from a sample without any culture isolation of microorganisms. The metagenomic approach is relatively quick that gives high-quality information in comparison to culture-based techniques (Hendriksen et al. 2019).

The metagenomic can be performed in two ways namely; 16S metagenomic sequencing and whole-metagenome sequencing (WMS). 16S metagenomic targets an amplicon of a small variable segment of a highly conserved 16S RNA gene present in all bacterial community. 16S metagenomic gives an insight on the possible microbial taxa and its relative abundance in a sample. On the other hand, in whole-metagenome sequencing approach, the entire genomic DNA is fragmented and sequenced without any amplification. The relative abundance of taxa and known functional and resistance genes in a sample can be determined by comparing the fragmented shotgun reads to available databases of known functional and resistance genes.

Other than the above two approaches, metagenomics has a longitudinal metagenomic approaches appropriate for studying the issues of burden and build-up of antibiotic resistance. Longitudinal metagenomic directly focuses a change in microbial resistance in a sample taken from a patient during a treatment course thus helps alleviating the emergence and transmission of resistance.

Therefore, the metagenomic approaches are useful for monitoring the antimicrobial resistance organisms and resistance genes using short-read sequencing that quantifies thousands of transmissible resistance genes in a single sample without microbial culture (Sukhum et al. 2019). It provides more accurate information on the presence of microbial taxa, pathogenesis and virulence. The metagenomic data would be useful for analyses of novel genes of interest. Owing to its direct application on samples from healthy or clinical cases, and on samples from the potential reservoir, metagenomics outstands as a tool for a single-point surveillance of antimicrobial resistance allowing identification of all resistance genes and their context in all reservoirs.

12.11 Use of Comprehensive Antibiotic Resistance Database (CARD)

CARD is basically a data organizing software system that provides high-quality reference records and achieved genomic sequence data within a defined vocabulary. For the research in resistome and genome-based antimicrobial resistance prediction,

the CARD biocuration team has created and included the Resistance Gene Identifier (RGI) software in the Antibiotic Resistance Ontology (ARO) system for the hindrance-free interaction with software development initiatives (Brian et al. 2020). The use of CARD was popularized in 2017 as a consequence of ease of curating exhaustive reference sequence, modifying ontological framework, ability to curate over 500 extra microbial resistance models, to facilitate the development of innovative classification paradigm and the expansion of analytical tools.

Recently there is an addition of a new module called "Resistomes and Variations" in CARD system which helps to analyze the in silico prediction of resistance variants from over 82 pathogens and one lakh genomes from the database. The inclusion of module on resistance variations has enabled the summarization of the expected resistance using the data in CARD. It has allowed identifying the trends in AMR mobility, understanding of the previously unexplained and novel resistance variants in microbes.

References

Abdelbary M, Basset P, Blanc D et al (2017) In: Tibayrenc M (ed) Genetics and evolution of infectious diseases: the evolution and dynamics of methicillin-resistant *Staphylococcus aureus*, 2nd edn. Elsevier, pp 553–572. ISBN: 9780127999425

Allard MW, Stevens EL, Brown EW (2019) All for one and one for all the true potential of whole-genome sequencing. Lancet Infect Dis 19:683–684

Antipov D, Hartwick N, Shen M et al (2016) Plasmid spades: assembling plasmids from whole-genome sequencing data. Bioinformatics 32:3380–3387

Biswas S, Raoult D, Rolain JM (2008) A bioinformatics approach to understanding antibiotic resistance in intracellular bacteria through whole-genome analysis. Int J Antimicrob Agents 32: 207–220

Blundell TL, Sibanda BL, Montalvao RW et al (2006) Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. Philos Trans R Soc B 361:413–423

Bradley P, Gordon NC, Walker TM et al (2015) Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *mycobacterium tuberculosis*. Nat Commun 6:10063. https://doi.org/10.1038/ncomms10063

Brian PA, Amogelang RR, Tammy TYL et al (2020) Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 2019(48):D1

Brown NA, Aisner DL, Oxnard GR (2018) Precision medicine in non-small cell lung cancer: current standards in pathology and biomarker interpretation. Am Soc Clin Oncol Educ book 38: 708–715

Carattoli A, Villa L, Feudi C et al (2017) Novel plasmid-mediated colistin resistance mcr-4 gene in Salmonella and Escherichia coli. Italy 2013, Spain and Belgium, 2015 to 2016. Eur Secur 22(31):30589. (Epub ahead of print)

Centers for Disease Control and Prevention (2013) Antibiotic resistance threats in the United States 2013. http://www.cdc.gov/drugresistance/threat-report-2013/. Accessed Apr 2022

Chen L, Yang J, Yu J et al (2005) VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res 33:325–328

Clausen PT, Zankari E, Aarestrup FM (2016) Benchmarking of methods for identification of antimicrobial resistance genes in bacterial whole-genome data. J Antimicrob Chemother 71: 2484–2488. https://doi.org/10.1093/jac/dkw184

Doyle RM, Burgess C, Williams R et al (2018) Direct whole-genome sequencing of sputum accurately identifies drug resistant mycobacterium tuberculosis faster than MGIT culture sequencing. J Clin Microbiol 56:e00666–e00618

Economou V, Gousia P (2015) Agriculture and food animals as a source of antimicrobial-resistant bacteria. Infect Drug Resist 8:49. https://doi.org/10.2147/IDR.S55778

Ellington MJ, Ekelund O, Aarestrup FM et al (2017) The role of whole-genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST subcommittee. Clin Microbiol Infect 23:2–22. https://doi.org/10.1016/j.cmi.11.012

Feijao P, Yao HT, Fornika D et al (2018) MentaLiST–a fast MLST caller for large MLST schemes. Microbial Genom 4:e000146. https://doi.org/10.1099/mgen.0.000146

Flandrois JP, Lina G, Dumitrescu O (2014) MUBII-TB-DB: a database of mutations associated with antibiotic resistance in mycobacterium tuberculosis. BMC Bioinformatics 15:107. https://doi.org/10.1186/1471-2105-15-107

Frost LS, Leplae R, Summers AO et al (2005) Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol 3:722–732

Gardy JL, Loman NJ (2018) Towards a genomics-informed, real-time, global pathogen surveillance system. Nat Rev Genet 19:9–20. https://doi.org/10.1038/nrg.2017.88

Gelband H, Miller-Petrie M, Pant S et al (2015) State of the World's antibiotics, 2015. Center for Disease Dynamics, Economics, and Policy, Washington, DC; http://cddep.org/publications/state_worlds_antibiotics_2015#sthash.l18BFUu2.dpbs

Gibson MK, Forsberg KJ, Dantas G (2015) Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. ISME J 9:207–216. https://doi.org/10.1038/ismej.2014.106

Giordano F, Aigrain L, Quail MA et al (2017) De novo yeast genome assemblies from MinION, PacBio and MiSeq platforms. Sci Rep 7:3935. https://doi.org/10.1038/s41598-017-03996-z

Gong J, Pan K, Fakih M et al (2018) Value-based genomics. Oncotarget 9(21):15792–15,815

Gonzaga-Jauregui C, Lupski JR, Gibbs RA (2012) Human genome sequencing in health and disease. Annu Rev Med 63:35–61

Gupta SK, Padmanabhan BR, Diene SM et al (2014) ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 58:212–220. https://doi.org/10.1128/AAC.01310-13

Gwinn M, Maccannell D, Armstrong GL (2019) Next generation sequencing of infectious pathogens. JAMA 321:893–894

Hendriksen RS, Munk P, Njage P et al (2019) Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. Nat Commun 10:1124–08853

Jesus TF, Ribeiro-Gonçalves B, Silva DN et al (2019) Plasmid ATLAS: plasmid visual analytics and identification in high-throughput sequencing data. Nucleic Acids Res 47:D188–D194

Jia B, Raphenya AR, Alcock B et al (2017) CARD. 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res 45:D566–D573. https://doi.org/10.1093/nar/gkw1004

Karp BE, Tate H, Plumblee JR et al (2017) National antimicrobial resistance monitoring system: two decades of advancing public health through integrated surveillance of antimicrobial resistance. Foodborne Pathog Dis 14:545–557. https://doi.org/10.1089/fpd.2017.2283

Katsanis SH, Katsanis N (2013) Molecular genetic testing and the future of clinical genomics. Nat Rev Genet 14(6):415–426

Kos VN, Déraspe M, McLaughlin RE et al (2015) The resistome of Pseudomonas aeruginosa in relationship to phenotypic susceptibility. Antimicrob Agents Chemother 59:427–436. https://doi.org/10.1128/AAC.03954

Lakin SM, Dean C, Noyes NR et al (2017) MEGARes: an antimicrobial resistance database for high-throughput sequencing. Nucleic Acids Res 45:574–580. https://doi.org/10.1093/nar/gkw1009

Lanza VF, Baquero F, Martinez JL et al (2018) In-depth resistome analysis by targeted metagenomics. Microbiome 6(11):0387

Liu B, Pop MB (2009) Antibiotic resistance genes database. Nucleic Acids Res 37:D443–D447. https://doi.org/10.1093/nar/gkn656

Lu H, Giordano F, Ning Z (2016) Oxford Nanopore MinION sequencing and genome assembly. Genomics Proteomics Bioinform 14:265–279. https://doi.org/10.1016/j.gpb.2016.05.004

Luscombe NM, Greenbaum D, Gerstein M (2001) What is bioinformatics? A proposed definition and overview of the field. Methods Inf Med 40:346–358

Mason A, Foster D, Bradley P et al (2018) Accuracy of different bioinformatics methods in detecting antibiotic resistance and virulence factors from Staphylococcus aureus whole-genome sequences. J Clin Microbiol 56:e01815–e01817

Mazzarotto F, Olivotto I, Walsh R (2020) Advantages and perils of clinical whole-exome and whole-genome sequencing in cardiomyopathy. Cardiovasc Drugs Ther 34(2):241–253

McArthur AG, Waglechne RN, Nizam F et al (2013) The comprehensive antibiotic resistance database. Antimicrob Agents Chemother 57:3348–3357. https://doi.org/10.1128/AAC.00419-13

Meienberg J, Bruggmann R, Oexle K et al (2016) Clinical sequencing: is WGS the better WES? Hum Genet 135(3):359–362

Miotto P, Tessema B, Tagliani E et al (2017) A standardised method for interpreting the association between mutations and phenotypic drug resistance in mycobacterium tuberculosis. Eur Respir J 50:1701354. https://doi.org/10.1183/13993003.01354-2017

Ndagi U, Falaki AA, Abdullahi M et al (2020) Antibiotic resistance: bioinformatics-based understanding as a functional strategy for drug design. RSC Adv 10(31):18451–18468

Ng PC, Kirkness EF (2010) Whole-genome sequencing. Methods Mol Biol 628:215–226

O'Leary NA, Wright MW, Brister JR et al (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion and functional annotation. Nucleic Acids Res 44:D733–D745

Pettengill EA, Pettengill JB, Binet R (2014) Phylogenetic analyses of Shigella and Enteroinvasive Escherichia coli for the identification of molecular epidemiological markers: whole-genome comparative analysis does not support distinct genera designation. Front Microbiol 6:1573. https://doi.org/10.3389/fmicb.2015.01573

Quainoo S, Coolen JPM, van Hijum SAFT et al (2017) Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. Clin Microbiol Rev 4:1015–1063

Rapin N, Lund O, Bernaschi M (2010) Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. PLoS One 4:e9862

Rhoads A, Au KF (2015) PacBio sequencing and its applications. Genomics, Proteomics Bioinformatics 13:278–289. https://doi.org/10.1016/j.gpb.2015.08.002

Robertson J, Nash JHE (2018) MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. Microb Genom 4:1–7

Rozov R, Kav AB, Bogumil D (2017) Recycler: an algorithm for detecting plasmids from de novo assembly graphs. Bioinformatics 33:475–482

Sandgren A, Strong M, Muthukrishnan P et al (2009) Tuberculosis drug resistance mutation database. PLoS Med 6:e2. https://doi.org/10.1371/journal.pmed.1000002

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci 74:5463–5467

Stevens EL, Timme R, Brown EW et al (2017) The public health impact of a publically available, environmental database of microbial genomes. Front Microbiol 8:808

Sukhum KV, Diorio-Toth L, Dantas G (2019) Genomic and metagenomic approaches for predictive surveillance of emerging pathogens and antibiotic resistance. Clin Pharmacol Ther 106(3):512–524. https://doi.org/10.1002/cpt.1535

Vincent BM, Lancaster AK, Scherz-Shouval R et al (2013) Fitness trade-offs restrict the evolution of resistance to amphotericin B. PLoS Biol 11(10):e1001692

Votintseva AA, Bradley P, Pankhurst L et al (2017) Same-day diagnostic and surveillance data for tuberculosis via whole-genome sequencing of direct respiratory samples. J Clin Microbiol 55: 1285–1298. https://doi.org/10.1128/JCM.02483-16

Wang F, He X, Huang B, Chen P et al (2015) Detection and analysis of resistance mutations of hepatitis B virus. Int J Clin Exp Med 8:9630–9639

Xavier BB, Lammens C, Ruhal R et al (2016) Identification of a novel plasmid-mediated colistin-resistance gene, mcr-2, in Escherichia coli, Belgium, June 2016. Euro Surveill 21(27):30280. https://doi.org/10.2807/1560-7917.ES.2016.21.27.30280

Zankari E, Hasman H, Cosentino S et al (2012) Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67:2640–2644. https://doi.org/10.1093/jac/dks261

Zankari E, Allesøe R, Joensen KG et al (2017) PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. J Antimicrob Chemother 72:2764–2768. https://doi.org/10.1093/jac/dkx217

Zetner A, Cabral J, Mataseje L et al (2017) Plasmid Proler: comparative analysis of plasmid content in WGS data. BioRxiv 121:350

Zhang X, Payne M, Lan R (2019) In silico identification of Serovar-specific genes for salmonella serotyping. Front Microbiol 10:835

Zhao EY, Jones M, Jones SJM (2019) Whole-genome sequencing in cancer. Cold Spring Harb Perspect Med 9(3):a03457

Zhou F, Xu Y (2010) cBar: a computer program to distinguish plasmid-derived from chromosome-derived sequence fragments in metagenomics data. Bioinformatics 26(16):2051–2052. https://doi.org/10.1093/bioinformatics/btq299

Chapter 13
Microbial Genomics and Modulation in Ruminants: An Environmental Perspective with Special Reference to Methane Migration

Sarmistha Debbarma, Jupi Talukdar, Anindita Sarma, Prabhakar Maurya, Dipak Deka, and Luit Barkalita

Abstract Rumen is considered the primary source for methane production, and its microbiome consisting of bacteria, yeast and protozoa help in methanogenesis by utilizing various biochemical pathways. Methane production has been an environmental concern for many years. Methane is far more effective amongst the greenhouse gases in trapping heat and plays a major role on the environment in respect to climate change. With accordance to the microbial genomics, their role in methane production and modification of diet in ruminants, researchers have been working on various methods to apply in the field of methane migration. Functional metagenomics and ruminomics are being studied and applied in the process of mitigation of methane production by ruminants. Here, we intend to provide an update on the microbial genomics and modulation in ruminants, and how the reviewed methods could help in preventing methane emissions.

Keywords Genomic selection · Metabolomics · Metagenomics · Microbial genomics · Methane migration · Methanogenesis · Functional metagenomics · Ruminants · Ruminomics

S. Debbarma
Department of Animal Resources Development, Government of Tripura, Agartala, Tripura, India

J. Talukdar
Mahisah Perspicientia LLP, Guwahati, Assam, India

A. Sarma · D. Deka · L. Barkalita (✉)
Department of Animal Biotechnology, Faculty of Veterinary Sciences, Assam Agricultural University, Guwahati, Assam, India
e-mail: dipk.deka@aau.ac.in; luit.barkalita@aau.ac.in

P. Maurya
CEHTRA (Fr) (India Office), New Delhi, India

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_13

13.1 Introduction

The current nutritional and food security demands has led to an increased global consumption of livestock products and it is expected to double by 2050 with many factors contributing to it such as better standards of living and increased population growth to name a few. Climate change has impacted all sectors of food production including livestock production and to combat the effects on climate change of livestock production and vice versa, sector-specific adaptation and mitigation strategies are required for sustainable production and growth of livestock sector (Rojas-Downing et al. 2017). The environmental footprint of agriculture and animal husbandry is largely associated with methane emissions due to ruminants (Tapio et al. 2017). Out of all the human-caused greenhouse gas emissions, 14.5% emissions come from livestock supply chains alone (FAO 2016). Robust research in the domain of livestock production and climate change sector has established the relationship between rumen microbial processes and the environment in terms of methane production, and an understanding of these processes with specific functions of rumen microbiomes can help towards deciding improved managemental practices related to feeding and also help to maintain a low environmental impact (Bekele et al. 2021).

For decades, research has focussed on deciphering the relationship between the inhabitant microbiome and the host ruminants. Though initial research studies aimed on identification and classification of the types of microbiotas and understanding the functional properties exerted by them, especially towards the energy requirements (Jami et al. 2014; Denman et al. 2018). The cellulosic plant materials are not directly fit for human consumption, and due to the inherent capability of rumen microbiome ecosystem to coevolve with the physiological system of the ruminants, conversion of fodder to high-protein products is possible. Research has been able to help in identification of a large number of microbial phylotypes using various molecular approaches employing culture-independent approaches and using single genome sequencing, it has been possible to accumulate a large database of genomic information on rumen microbiome ecosystem. However, due to constraints in cultivation of many of such microbiomes in laboratory conditions, a data gap remains in the understanding of functional characteristics of these (Li et al. 2018).

With the advent of new knowledge and access to improved technologies as well as better technical knowhow, the research has shifted from studies on single strains of rumen microbiomes to 'total' microbiota through large-scale sequencing of nucleic acids using techniques such as metagenomics, meta transcriptomics, proteomics and metabolomics (Jami et al. 2014; Denman et al. 2018). To develop a better understanding of rumen fermentation mechanisms which are vital for incorporating changes towards improvement in nutrient utilization and in general productivity of ruminants, historical data as well as advanced techniques such as metagenomics, meta transcriptomics and other bioinformatics methods can be of utmost importance (Li et al. 2018). Methane mitigation based on small-molecule inhibitor and vaccines in ruminants can be improved using genome sequencing data of ruminal microbiota

as well as interpretation of large-scale nucleic sequencing studies. Genome sequencing also paves the way for other methane mitigation strategies such as advances in genetic selection and breeding of animals which are genetically inclined to be low emitters and modification of ruminal microbiota ecosystem, thus tweaking the microbial fermentation dynamics (Leahy et al. 2013).

Livestock production is expected to be affected by climate change as variability in many aspects of livestock management and production is inevitable. The global water consumption by animals is expected to increase up to three times, and an increase in the requirement of agricultural lands due to expected growth in production by 70%, thus leading to increased concerns of food security is a possibility, since livestock fodder utilizes a good one-third of global cereal production. Therefore, for successful transition of livestock production towards sustainability, there is a need to incorporate such measures of management, production, adaptation or mitigation which are streamlined with local practices and are feasible with the particular set of resources. These have to be supported by policies and regulations in consensus with the scientific rationale for combating effects of climate change and promoting mitigation and/or adaptation measures (Rojas-Downing et al. 2017).

13.2 Biochemical Aspects of Rumination

Ruminants have a very distinct microbiome in the rumen, with respect to biochemical relationship between the feed and microorganisms. These microorganisms help in degradation by way of fermentation of feed making it feasible for the hosts to utilize the final products of fermentation for their nutrition as well as source of energy. Being the largest compartment, the rumen is lined with numerous papillae which resemble carpet-like structures. These papillae are known to extend from the rumen wall which increases the surface area and thus enhancing the capability of absorption. Observations have been made that more the population of microbiome in the rumen, there was an increase in the length and width of the rumen papillae.

The rumen is considered the 'warehouse' for rumination and production of methane (CH_4) gas. Various biochemical pathways are involved in the process of methanogenesis. Being an anaerobic ecosystem, the rumen is crucial for conversion of carbohydrates (CHOs) to short-chain volatile fatty acids (VFAs). The cellulolytic bacteria in the ruminal microbiome helps in breakage of the β $(1 \rightarrow 4)$ linkages of carbohydrates present in the plant cell wall, and these microorganisms further utilize the liberated hexoses and pentoses to provide the host with energy. These VFAs are absorbed by the host via process of nonionic diffusion across the fore stomach epithelium and these acids are utilized in the metabolism of energy and protein. During the process of fermentation, dihydrogen is produced which is mostly utilized by the archaea (methanogenic) present in the rumen and they reduce the carbon dioxide ($CO_{2)}$ gas to methane (CH_4) (Hungate 1967). Fermentation in rumen occurs with the help of intracellular and extracellular flows of metabolic dihydrogen gas, with respect to the production of VFAs, incorporation of dihydrogen gas into the

competing biochemical pathways and interspecies transfer of dihydrogen gas. Acetate and butyrate, produced via glycolysis, are associated with production of hydrogen ions and propionate is responsible for incorporation of hydrogen ions into the methanogenesis reaction. The hydrogen released due to production of acetate and butyrate is utilized by the methanogens in the rumen to reduce the carbon dioxide ($CO_{2)}$ gas to methane (CH_4) (Janssen 2010). When methane (CH_4) is released to the atmosphere from rumen, there is also a loss of energy for ruminants viz. about 2–12% of a cow's gross energy intake (Eckard et al. 2010; Martin et al. 2010; Johnson and Johnson 1995; Yan et al. 2010). With respect to environmental perspective, researchers have been working on various methane mitigation strategies to reduce the enteric methanogenesis itself as well as its release to the environment through modification of microbiome of the rumen and/or by modifying the diet and observing its effects on methanogenesis.

13.3 Role of Rumen Microbiome in Methane Emission

Methane (CH_4) exists in gaseous form and is a natural byproduct of enteric fermentation in ruminants (Fig. 13.1) contributing 6% of global emissions of methane from anthropogenic sources (Difford et al. 2018). As per present understanding, it is now evident that ruminant livestock make a significant contribution towards the emission of methane and some specificities of ruminal microbiota are associated with the methane phenotypes which can be low as well as high (Tapio et al. 2017).

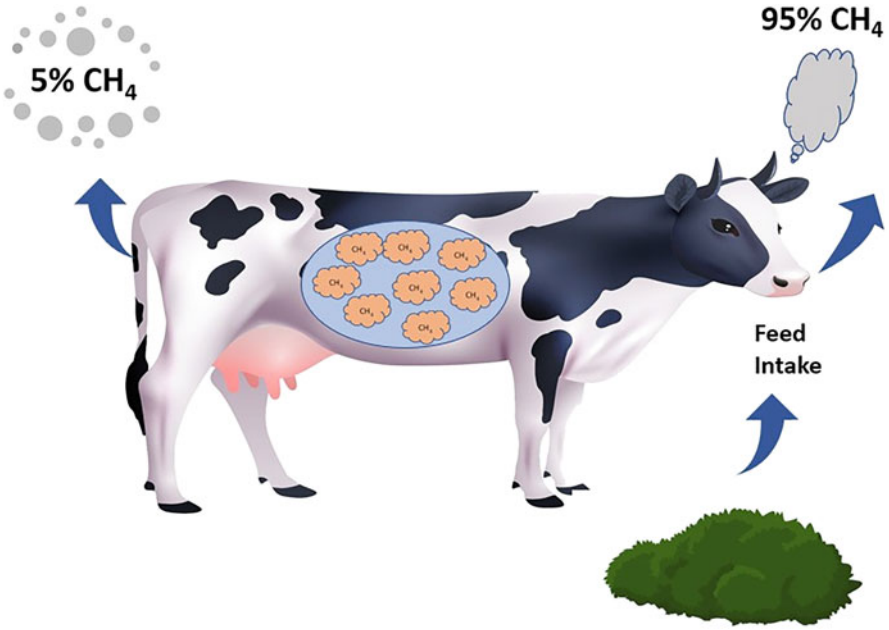

Fig. 13.1 Formation and emission of enteric methane during a cow's digestive process. (Source: The Author)

Ruminants comprising chiefly of cattle as well as giraffes, deer, antelopes, sheep, goats, etc., produce humongous amounts of methane, about 110 million metric tonnes per annum and being a GHG, it has a tremendous effect on climate change. The main sources which contribute to these emissions are feed production, feed processing and methane generated due to ruminant's digestion. The positive news is that adoption of established impactful global practices and interventions relating to animal feeding and management could help to globally cut down emissions from the livestock sector of greenhouse gases up to 30% and thus be more resilient. The Food and Agriculture Organization (FAO) is a strategic partner to many initiatives which address the impacts due to global climate change on the livestock sector as well as which aim at reducing the greenhouse gas emissions by ruminants and other domestic animals (FAO 2016).

The development of appropriate mitigation strategies towards reducing methane emissions from livestock incites interest in both scientific research and environmental-concerned communities. Being the sole producers of ruminant methane, methanogens are the main target of methane abatement strategies. It is very important to understand the relationship between methanogens and the rest of ruminal microflora for developing mitigating strategies for methane in ruminants. Methane abatement strategies also target other rumen microbial community which produces substrates necessary for methanogenesis (Leahy et al. 2013). Recent advances such as functional metagenomics screening technology can greatly enhance our understanding of the rumen microbial process (Bekele et al. 2021). Metagenomics and meta transcriptomics can provide high-resolution taxonomic and functional information by sequencing of nucleic acids from different environmental samples, thus capturing whole genome and transcriptome repertoire of microorganisms (Li et al. 2018).

13.4 Factors and Effects of Enteric Methane Emission

13.4.1 Environmental Effects of Methane Emission

Enteric methane is a Short-Lived Climate Pollutant (SLCP) but with a comparatively much shorter atmospheric half-life of 12 years in contrast to CO_2 which has an atmospheric half-life of >100 years (IPCC 2014). Methane has the ability to trap up to 84 times more heat than carbon dioxide for up to two decades once it is released into the atmosphere, and when compared with carbon dioxide (per kg) for up to a century, methane has 34 times greater warming effects. Thus, controlling the enteric methane emissions by way of reduction can bring near-term reductions in warming and sustained reductions can possibly limit peak warming as well.

Globally, the annual production of enteric methane by ruminant livestock is about 3.3 Gt CO_2 equivalents. Out of this, cattle account for 77% (2.5 Gt), buffalo for 13% (0.43 Gt) and small ruminants (sheep and goats) for the remainder (0.31 Gt) (FAO 2022) (Fig. 13.2). It seems a simple thought that addressing the issue of enteric

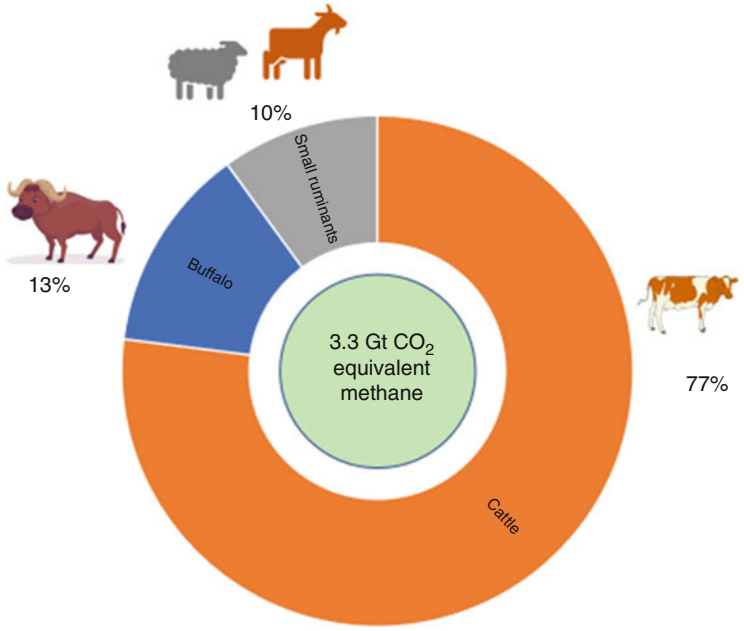

Fig. 13.2 Contribution of enteric methane to global GHGs (as CO_2 equivalents). (Source: The Author)

methane completely can deliver instant solutions towards climate change mitigation but in practical scenario, it is not possible to fully control or avoid the production of enteric methane. Moreover, it is not a short-term possibility to even implement the already available methods and strategies towards mitigating enteric methane emissions, as there is an increased demand for livestock-based products and it is expected to rise over the years to come. A more pragmatic approach will be to focus on emission intensity which is reducing emissions per unit of product. This is possible mostly by improving the production efficiency by implementing known methods, techniques, strategies and technologies aiming towards higher yield per unit of animal and per unit of feed provided (FAO 2022).

Methane (CH_4) has secondary environmental effects as well, as it is the main contributor in the formation of ozone at the ground level. Ozone is a hazardous greenhouse gas classified as a toxic air pollutant, and it is considered to be responsible for about one million premature deaths per year. Since preindustrial times, nearly 30% of all global warming can be credited to methane emissions. According to United States National Oceanic and Atmospheric Administration, even as CO_2 emissions were reduced during the pandemic-related lockdowns of 2020, atmospheric CH_4 levels were increased (UNEP 2021).

13.4.2 Factors Affecting Ruminant Methane Modulation

Methane production by ruminants is modulated due to various factors such as feed intake and composition as well as the animal's individual genetic makeup and composition of ruminal microbiota. The abundance of certain bacteria and archaea are up to limited extent influenced by the genetic makeup of host and some taxa contribute to a greater extent towards methane production than others (Difford et al. 2018). The environmental and genetic factors such as breed, genotype, fodder, management and physiological status of the animal also influence the enteric methane production. The microbiological parameters of enteric methanogenesis have been extensively researched, but for having an accurate understanding of important differences in the host genotype and microbial genome with the methods of interaction, further research is needed (Asselstine et al. 2021). A study by Difford et al. (2018) demonstrated that variations in methane production are not likely modulated by genetic changes on the rumen microbiota. Thus, it is a possibility to target rumen microbiota and host genome independently, by selective breeding of low methane-producing ruminants and parallelly experimenting with methods targeting changes in the rumen microbiota to reduce methane production and emissions in the ruminants.

13.5 Mitigation Strategies for Ruminant Methane Emission

13.5.1 Traditional Approaches for Enteric Methane Mitigation

Through years of scientific research, many methods have been developed for mitigating enteric methane production. These methods of enteric methane mitigation have been variedly described but can be broadly divided into three broad categories: Dietary manipulation, Rumen modulation and Animal management (Fig. 13.3) (Knapp et al. 2014; Reisinger et al. 2021; Black et al. 2021; Tamburini 2020; Arndt et al. 2021; Martin et al. 2010). Dietary manipulation strategies focus on the managemental and nutritional aspects related to feed for increasing animal productivity and feed efficiency. Rumen modulation strategies focus on feeding of such substances which can directly or indirectly reduce methane production or on using biological controls such as defaunation, bacteriocins, bacteriophages and immunization which aim to reduce methanogens itself. Managemental strategies focus on improving productivity of animals through genetic manipulation as well as by incorporating changes in managemental practices such as improvements in nutrient utilization towards production, improving efficiency of feed and reducing methane emission per unit of product (meat or milk).

Fig. 13.3 Traditional approaches for enteric methane mitigation. (Source: The Author)

13.5.2 Gut Microbiome Modulation in Ruminants

Being a very significant greenhouse gas, enteric methane gas (CH_4) production has been a social and environmental concern as well as a channel for the loss of energy in ruminants. Keeping in mind the effects of methane gas in the environment, there have been numerous developments of CH_4 mitigation strategies in the research field, with respect to ruminant nutrition (Patra et al. 2017). Gut microbiome modulation is one of the strategies which is considered as a process for methane mitigation and was studied as early as in the year 1960s when research on H_2-utilization pathway was conducted (Baldwin et al. 1963). The research findings mentioned that rumen microbiota converts pyruvate to propionate and simultaneously, there was reduction of protons (H^+) to H_2 (main precursor in methane production). Due to an increase in propionate formation, there was stoichiometric decrease in methane production.

The acrylate pathway was studied by Russell and Wallace (1997), which is a critical propionate-producing pathway of the rumen and is utilized by *Megasphaera elsdenii* in the presence of lactate. With the help of the acrylate pathway, *M. elsdenii* helped in reducing methane production and it was found that cows supplemented with a *M. elsdenii* DFM (direct fed microbials) showed changes in pattern of rumen fermentation by increasing propionate formation along with improving energy balance and productivity of animals (Henning et al. 2010; Aikman et al. 2011). Krehbiel and co-workers (2003) suggested that use of DFM is a possible way to reduce methane emission. In the year 2008, Chaucheyras-Durand and Group (2008) worked with dry yeasts as DFM and they reported three beneficial effects of yeast DFM: an improvement in maturity of rumen by increasing establishment of microbiomes, stabilizing of the ruminal pH and an enhanced degradation of fibre. McAllister and Group (2011) stated that the function of the dry yeast, as mentioned above, might have reduced methane production by removal of oxygen by the yeast or due to the presence of micro-nutrients in the yeast, which helped to lower the production of methane. Cellulolytic bacterium *F. succinogenes* can be used as DFM due to its ability to produce succinate, which leads to increased propionate production with less H_2 formation. Following this theory, Chaucheyras-Durand and Group (2010) inoculated *F. succinogenes* culture as DFM into gnotobiotically reared lambs and observed significant decrease in methanogenesis in vitro. Newbold and co-workers (2015) worked with ciliate protozoa to learn about the protozoa's function in the rumen. They concluded that although rumen protozoa stay as inhabitants, they are not preferred as essential or important in the rumen microbiome. They proposed that the elimination of protozoa from the rumen can lead to enhanced nutrient flow to the small intestine as well as decrease rumen ammonia production and methanogenesis. Wang et al. (2018) studied the shifts of hydrogen metabolism during the methanogenesis process with changes in the diet. They observed that dietary change of forage fibre with non-forage fibre in the diet led to a substitution of *Firmicutes* by *Bacteroidetes* and of *Methanobrevibacter* by *Methanomassiliicoccus,* which resulted in an enhanced shift towards propionate production due to hydrogen flow.

13.5.3 Omics Approaches for Enteric Methane Mitigation

The application of –omics-based methods such as metabolomics and metagenomics can be of critical importance in ascertaining information which can not only help in deciphering genetic basis of methanogenesis but also develop mitigation measures towards reducing enteric methane emissions (Fig. 13.4). Omics approaches can aid in covering the data gaps around genetic basis of interaction between host genotype and ruminal microflora. Using systems biology approaches, different -omics technologies can be used to identify genetic markers of methane production in ruminants which can be of prime importance in selective breeding of low methane-producing animals. Due to cost effectiveness, there exist some challenges in employing –omics-based methods towards reducing methane emissions. Systems biology approach which can help to integrate as well as evaluate different levels of biological information can provide specific insights towards deciphering genetic basis of methanogenesis as well as aid in developing methods for its mitigation (Asselstine et al. 2021). Systems biology methods follow a top-down and bottom-up approach to understand various levels of biological information and its influence on a system (Shahzad and Loor 2012).

13.5.3.1 Metagenomic Profiling and Methane Emissions in Ruminants

The application of metagenomics in microbiome analysis can help to characterize metagenome of ruminal microbiome and its association with performance traits (Carberry et al. 2014a, b; Roehe et al. 2016; Kamke et al. 2016; Wang et al. 2017; Hess et al. 2020). The main methods used for measuring metagenomes of rumen are:

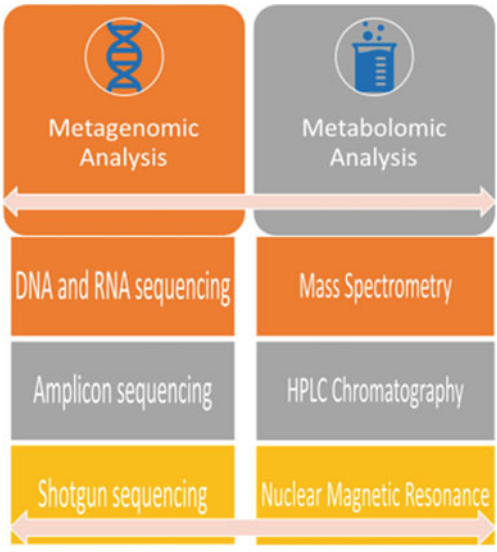

Fig. 13.4 Omics approaches for enteric methane mitigation. (Source: The Author)

DNA sequencing, RNA sequencing, Amplicon Sequencing and Shotgun Sequencing. There are benefits as well as limitations of these methods. Benefits: Traditional DNA sequencing is targeted and specific and provides long continuous reads; RNA sequencing requires low sequencing depth and is highly sensitive and reproducible; Amplicon sequencing can characterize target sequences, i.e., 16 s rRNA, 18 s rRNA genes, OTUs; and Shotgun sequencing is high-throughput and provides ability of parallel sequence and obtain unbiased microbiome profiles. Limitations: Traditional DNA sequencing has amplification bias and is time consuming; RNA sequencing requires biological replicates and the accuracy depends on annotation quality of reference genome; Amplicon sequencing is low-thoroughput; and Shotgun sequencing requires high sequencing depth.

13.5.3.2 Metabolomic Profiling and Methane Emissions in Ruminants

Metabolomic profiling can be used as an essential tool in deciphering the mechanisms of biological processes in relation to genetics and environmental aspects in ruminants and can provide valuable inputs for applications in breeding or managemental aspects. The main methods used for measuring metagenomes of rumen are: Higher Performance Liquid Phase Chromatography (HPLC), Mass Spectrometry (MS) and Nuclear Magnetic Resonance (NMR). There are benefits as well as limitations of these methods. Benefits: MS has high selectivity and sensitivity for metabolite identification; HPLC can help to identify compounds in multi-component mixtures; and NMR requires minimal sample preparation. Limitations: In MS, the preparation steps can damage metabolites; HPLC is not cost-effective; and in NMR, individual metabolite identification in complex mixtures can be challenging due to clustering effects.

13.5.3.3 Integration of '-Omics' and Functional Analyses

The integration of different -omics technologies being high-throughput in nature and functional analyses using a systems biology approach can help to decipher complex information on functional genes and their influence on the regulation of complex trait phenotypes including methane (Cánovas et al. 2014a, b; Cánovas 2016; Fonseca et al. 2018). In addition to metabolomics and metagenomics, transcriptomics is a highly application-based -omics technology which can evaluate the transcriptome at a high-throughput level. Transcriptomics helps to ascertain data on total expressed RNA in a cell or tissue at a specific time stamp, and it aids in measuring gene expression levels, as well as detection of structural variants such as single-nucleotide polymorphisms (SNP), insertions and deletions and identification of differentially expressed splice variants in the entire transcriptome (Wang et al. 2009).

Such genetic basis of information using -omics technologies and systems biology approaches on specific genetic traits can help in improved breeding programs,

thereby improving efficiency of production (Cánovas 2016; Suravajhala et al. 2016; Fleming et al. 2018). Such insights into data can be used at the industrial level to develop an insight into biological and genetic information which can be utilized to improve the understanding of performance traits resulting in selection accuracy, reducing inter-generation intervals as well as increased rate of genetic improvement of traits (Cánovas 2016; Suravajhala et al. 2016).

13.6 Current Knowledge and Way Forward

Livestock are an inherent part of agriculture and still form the major backbone of economy in many developing nations employing millions of people worldwide. There is a direct correlation between livestock production and climate change and vice versa. Climate change has major impacts on livestock keepers and on the ecosystem's goods and services on which they depend (OECD/FAO 2019). By the year 2050, the growth of population of humans is expected to be over 9 billion, and global consumption of meat is expected to rise by 73% (OECD/FAO 2019). Livestock contributes about 14.5% of the total anthropogenic-related GHG emissions and ruminants alone contribute about 5.8% by enteric methane emissions (Gerber et al. 2013; Difford et al. 2018). Multiple research studies stress on reducing the methane emissions as anthropogenic-related emissions of methane is a key area where mitigation strategies can deliver a significant impact (IPCC 2014; Gerber et al. 2013) as it is a potent greenhouse gas with a very high climate change potential up to ≈32 times of carbon dioxide (CO_2) (Holmes et al. 2013) but with a comparatively much shorter atmospheric half-life of 12 years in contrast to CO_2 which has an atmospheric half-life of >100 years (IPCC 2014). Therefore, reducing methane emissions from anthropogenic-related sources has been identified as a key area for mitigating climate change with immediate effects (IPCC 2014; Gerber et al. 2013).

Ruminants have a multichambered stomach and due to the presence of a distinct ruminal microbiome, they are capable of digesting a variety of high fibre feedstuffs. Methane is a natural byproduct formed on gastro-enteric fermentation of fibrous biomass of plants by the action of ruminal microbial enzymatic activity (Hill et al. 2016). Hydrogen and CO_2 are produced in the rumen by action of various bacteria, protozoa or fungi and on further action of archaea called methanogens, these are converted to methane and about 99% of methane is released in the breath by livestock through eructation and respiration (Janssen and Kirs 2008). The enteric methane emission is considered as an important physiological pathway for maintenance of ruminal pH and hydrogen balance as the optimal conditions are limited to a narrow range of partial pressure of H_2 and pH, wherein anaerobic fermentation by the rumen microbial community is possible. Metabolic hydrogen from anaerobic fermentation is converted into $H_2$by hydrogenase-expressing bacteria which is further converted to methane via methanogenesis (McAllister and Newbold 2008). Emitted methane has a caloric value and it results in losses of about 2–12% of a cow's gross energy intake (Johnson and Johnson 1995; Yan et al. 2010).

Consequently, cattle and such ruminants having an increased efficiency for digesting high fibre feedstuffs but seemingly reduced methane production would be the ideal choice for benefitting climate as well as maintaining a sustainable livestock production practice.

Collaborative international research studies have been targeting to understand the physiological mechanisms of rumen microbiota through various interventions such as chemical or biological feed additives, feed formulations and anti-methanogen vaccines (Moss et al. 2000), but the capability to rapidly adapt to changes in the substrate complex by ruminal microbiota renders only limited success by such interventions and the methane production again returns to pre-treatment levels (Hristov et al. 2013). There is scientific indication of host influence over rumen microbial composition as in rumen transplantation studies (transfaunation), rumen bacteria recovered within a short period of time to pre-transfaunation composition (Weimer et al. 2010). The host genotype of cattle was able to explain the changes associated with inter-animal differences in relation to methane production (Lassen and Løvendahl 2016; Lassen et al. 2016; Donoghue et al. 2016) and the effects on ruminal microflora on the methane production (Roehe et al. 2016). However, exact empirical evidence which can link the role of host's genotype with ruminal microflora and/or methane genesis is rather limited (Roehe et al. 2016).

A promising methodology to mitigate enteric methane production is through selecting genetically low-emitting cattle and other livestock, as not only it will be a sustainable initiative but its cumulative effect over subsequent generations can have long-term reductions of methane. If reduced enteric-methanogenesis is due to an anomaly with ruminal microbial population, there arises a possibility of risk in selection and thus affecting the very symbiotic relationship of coexistence of ruminants and ruminal microbial population. Thus, it is very pertinent to critically analyse the extent of relationship between the ruminal microbiota and host genetic makeup. The omics technologies have provided a pathway to enhance our understanding of species-specific functional knowledge with their genetic makeup. Such knowledge could help in developing an enhanced ability in articulating targeted interventions which can direct ruminal composition and its activity for better production and health (Denman et al. 2018).

13.7 Conclusion

In the complete ecosystem of livestock production, the major contributor towards the GHG emission is the feed production stage and the enteric methanogenesis is the highest contributor to greenhouse gas emission in the animal production. Both these are slated to rise in time-bound manner due to intensification of livestock production along with the demand. Thus, with an increase in population of livestock and unchanged feeding practices, the global levels of emissions due to production processes will be continually increased over time. To make a radical shift towards sustainable livestock production practices and processes, holistic approach of

assimilation of best practices of adaptation and mitigation strategies supported by conducive policy frameworks are required. Improvement of animal feed and feeding practices with a focus on improved nutrition and improving as well as upgrading genetic makeup of livestock are important factors in mitigating enteric methane production. The need of the hour is better quantitative and qualitative estimates of CO_2 and GHG emissions related to various activities in production, processes, logistics and managemental practices within the framework of livestock products, though a clear gap seems to be present in current usage of control, adaptation or mitigation measures for various geographical and managemental systems in different parts of world (Rojas-Downing et al. 2017).

To understand the functional and systemic relationship of nutritional, production and managemental practices in ruminants to enteric methane production, it is necessary to have a cataloguing of rumen microbiota and assignment of functional characteristics of these genes creating a reference set of ruminal microbial genomes which can help in devising impactful strategies for enteric methane mitigation (Turnbaugh et al. 2007). Such strategy will also help to hypotheses which can be tested via experiments to have a better grasp at understanding rumen microbiome and its interaction with biological and environmental factors (Morgavi et al. 2013). Climate change is directly related to livestock production and as a result to food security. Similarly, livestock, especially ruminants, are also responsible for climate change due to release of methane in the environment which is formed due to enteric methanogenesis in ruminants (Gerber et al. 2013).

Historical data assimilated through years of research along with existing resources and usage of omics technologies in an integrated way can improve our understanding as well as advance our knowledge of metabolome and metagenome of ruminants. Thus, to have a better understanding of genetic and metabolic pathways influencing regulation and variation of enteric methanogenesis in ruminants, an integrated approach assimilating the advances of information from metabolomics, high-throughput omics, systems biology and historical data is required (Asselstine et al. 2021).

References

Aikman PC, Henning PH, Humphries DJ, Horn CH (2011) Rumen pH and fermentation characteristics in dairy cows supplemented with *Megasphaeraelsdenii* NCIMB 41125 in early lactation. J Dairy Sci 94(6):2840–2849. https://doi.org/10.3168/jds.2010-3783

Asselstine V, Lam S, Miglior F, Brito LF, Sweett H, Guan L, Waters SM, Plastow G, Cánovas A (2021) The potential for mitigation of methane emissions in ruminants through the application of metagenomics, metabolomics, and other -OMICS technologies. J Anim Sci 99(10):skab193. https://doi.org/10.1093/jas/skab193

Arndt C, Hristov AN, Price WJ, McClelland SC, Pelaez AM, Cueva SF, Oh J, Bannink A, Bayat AR, Crompton LA et al (2021) Strategies to mitigate enteric methane emissions by ruminants - a way to approach the 2.0°C target. AgriRxiv pp53:61. https://doi.org/10.31220/agriRxiv.2021.00040

Baldwin RL, Wood WA, Emery RS (1963) Conversion of glucose-C_{14} to propionate by the rumen microbiota. J Bacteriol 85(6):1346–1349. https://doi.org/10.1128/jb.85.6.1346-1349.1963

Bekele W, Zegeye A, Simachew A, Assefa G (2021) Functional metagenomics from the rumen environment-a review. Adv Biosci Biotechnol 12(5):125–141. https://doi.org/10.4236/abb.2021.125009

Black JL, Davison TM, Box I (2021) Methane emissions from ruminants in Australia: mitigation potential and applicability of mitigation strategies. Animals (Basel) 11(4):951. https://doi.org/10.3390/ani11040951

Carberry CA, Kenny DA, Kelly AK, Waters SM (2014a) Quantitative analysis of ruminal methanogenic microbial populations in beef cattle divergent in phenotypic residual feed intake (RFI) offered contrasting diets. J Anim Sci Biotechnol 5(1):41. https://doi.org/10.1186/2049-1891-5-41

Carberry CA, Waters SM, Kenny DA, Creevey CJ (2014b) Rumen methanogenic genotypes differ in abundance according to host residual feed intake phenotype and diet type. Appl Environ Microbiol 80(2):586–594. https://doi.org/10.1128/AEM.03131-13

Cánovas A, Reverter A, DeAtley KL, Ashley RL, Colgrave ML, Fortes MR, Islas-Trejo A, Lehnert S, Porto-Neto L, Rincón G et al (2014a) Multi-tissue omics analyses reveal molecular regulatory networks for puberty in composite beef cattle. PLoS One 9(7):e102551. https://doi.org/10.1371/journal.pone.0102551

Cánovas A, Rincón G, Bevilacqua C, Islas-Trejo A, Brenaut P, Hovey RC, Boutinaud M, Morgenthaler C, VanKlompenberg MK, Martin P et al (2014b) Comparison of five different RNA sources to examine the lactating bovine mammary gland transcriptome using RNA-sequencing. Sci Rep 4:5297. https://doi.org/10.1038/srep05297

Cánovas A (2016) Looking ahead: applying new genomic technologies to accelerate genetic improvement in beef cattle. Ceiba 54(1):41–49. https://doi.org/10.5377/ceiba.v54i1.2776

Chaucheyras-Durand F, Walker ND, Bach A (2008) Effects of active dry yeasts on the rumen microbial ecosystem: past, present and future. Anim Feed Sci Technol 145:5–26. https://doi.org/10.1016/j.anifeedsci.2007.04.019

Chaucheyras-Durand F, Masseglia S, Fonty G, Forano E (2010) Influence of the composition of the cellulolytic flora on the development of hydrogenotrophic microorganisms, hydrogen utilization, and methane production in the rumens of gnotobiotically reared lambs. Appl Environ Microbiol 76(24):7931–7937. https://doi.org/10.1128/AEM.01784-10

Denman SE, Morgavi DP, McSweeney CS (2018) Review: the application of omics to rumen microbiota function. Animal 12(s2):s233–s245. https://doi.org/10.1017/S175173111800229X

Difford GF, Plichta DR, Løvendahl P, Lassen J, Noel SJ, Højberg O, Wright AG, Zhu Z, Kristensen L, Nielsen HB, Guldbrandtsen B, Sahana G (2018) Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. PLoS Genet 14(10):e1007580. https://doi.org/10.1371/journal.pgen.1007580

Donoghue KA, Bird-Gardiner T, Arthur PF, Herd RM, Hegarty RF (2016) Genetic and phenotypic variance and covariance components for methane emission and postweaning traits in Angus cattle. J Anim Sci 94(4):1438–1445. https://doi.org/10.2527/jas.2015-0065

Eckard RJ, Grainger C, De Klein CAM (2010) Options for the abatement of methane and nitrous oxide from ruminant production: a review. Livest Sci 130:47–56. https://doi.org/10.1016/j.livsci.2010.02.010

FAO (2016). Livestock and climate change. https://www.fao.org/3/i6345e/i6345e.pdf. Accessed 15 Feb 2022

FAO (2022). Reducing enteric methane for improving food security and livelihoods. https://www.fao.org/in-action/enteric-methane/background/why-is-enteric-methane-important/en/. Accessed 15 Feb 2022

Fonseca PAS, Id-Lahoucine S, Reverter A, Medrano JF, Fortes MS, Casellas J, Miglior F, Brito L, Carvalho MRS, Schenkel FS et al (2018) Combining multi-OMICs information to identify key-regulator genes for pleiotropic effect on fertility and production traits in beef cattle. PLoS One 13:e0205295. https://doi.org/10.1371/journal.pone.0205295

Fleming A, Abdalla EA, Maltecca C, Baes CF (2018) Invited review: reproductive and genomic technologies to optimize breeding strategies for genetic progress in dairy cattle. Arch Anim Breed 61:43–57. https://doi.org/10.5194/aab-61-43-2018

Gerber PJ, Steinfeld H, Henderson B, Mottet A, Opio C, Dijkman J, Falcucci A, Tempio G (2013) Tack` climate change through livestock – a global assessment of emissions and mitigation opportunities. Food and agriculture Organization of the United Nations (FAO), Rome. https://www.fao.org/3/i3437e/i3437e.pdf. E-ISBN: 978-92-5-107921-8 (PDF)

Henning PH, Horn CH, Leeuw KJ, Meissner HH, Hagg FM (2010) Effect of ruminal administration of the lactate-utilizing strain *Megasphaeraelsdenii* (me) NCIMB 41125 on abrupt or gradual transition from forage to concentrate diets. Anim Feed Sci Technol 157:20–29. https://doi.org/10.1016/j.anifeedsci.2010.02.002

Hess MK, Rowe SJ, Van Stijn TC, Henry HM, Hickey SM, Brauning R, McCulloch AF, Hess AS, Kirk MR, Kumar S et al (2020) A restriction enzyme reduced representation sequencing approach for low-cost, high-throughput metagenome profiling. PLoS One 15:e0219882. https://doi.org/10.1371/journal.pone.0219882

Holmes CD, Prather MJ, Søvde OA, Myhre G (2013) Future methane, hydroxyl, and their uncertainties: key climate and emission parameters for future predictions. Atmos Chem Phys 13(1):285–302. https://doi.org/10.5194/acp-13-285-2013

Hristov AN, Oh J, Firkins JL, Dijkstra J, Kebreab E, Waghorn G, Makkar HP, Adesogan AT, Yang W, Lee C, Gerber PJ, Henderson B, Tricarico JM (2013) Special topics--Mitigation of methane and nitrous oxide emissions from animal operations: I. A review of enteric methane mitigation options. J Anim Sci 91(11):5045–5069. https://doi.org/10.2527/jas.2013-6583

Hill J, McSweeney C, Wright A-DG, Bishop-Hurley G, Kalantar-zadeh K (2016) Measuring methane production from ruminants. Trends Biotechnol 34(1):26–35. https://doi.org/10.1016/j.tibtech.2015.10.004

Hungate RE (1967) Hydrogen as an intermediate in the rumen fermentation. Arch Mikrobiol 59(1):158–164. https://doi.org/10.1007/BF00406327

IPCC (2014) In: Core writing team, Pachauri RK, Meyer LA (eds) Climate change 2014: synthesis report. Contribution of working groups I, II and III to the fifth assessment report of the intergovernmental panel on climate change. IPCC, Geneva, Switzerland, p 151. https://archive.ipcc.ch/pdf/assessment-report/ar5/syr/SYR_AR5_FINAL_full_wcover.pdf. Accessed 15 Mar 2022

Jami E, White BA, Mizrahi I (2014) Potential role of the bovine rumen microbiome in modulating milk composition and feed efficiency. PLoS One 9(1):e85423. https://doi.org/10.1371/journal.pone.0085423

Janssen PH, Kirs M (2008) Structure of the archaeal community of the rumen. Appl Environ Microbiol 74(12):3619–3625. https://doi.org/10.1128/AEM.02812-07

Janssen PH (2010) Influence of hydrogen on rumen methane formation and fermentation balances through microbial growth kinetics and fermentation thermodynamics. Anim Feed Sci Technol 160:1–22. https://doi.org/10.1016/j.anifeedsci.2010.07.002

Johnson KA, Johnson DE (1995) Methane emissions from cattle. Anim Sci J 73(8):2483–2492. https://doi.org/10.2527/1995.7382483x

Knapp JR, Laur GL, Vadas PA, Weiss WP, Tricarico JM (2014) Invited review: enteric methane in dairy cattle production: quantifying the opportunities and impact of reducing emissions. J Dairy Sci 97(6):3231–3261. https://doi.org/10.3168/jds.2013-7234

Kamke J, Kittelmann S, Soni P, Li Y, Tavendale M, Ganesh S, Janssen PH, Shi W, Froula J, Rubin EM, Attwood GT (2016) Rumen metagenome and metatranscriptome analyses of low methane yield sheep reveals a Sharpea-enriched microbiome characterised by lactic acid formation and utilisation. Microbiome 4:56. https://doi.org/10.1186/s40168-016-0201-2

Krehbiel CR, Rust SR, Zhang G, Gilliland SE (2003) Bacterial direct-fed microbials in ruminant diets: performance response and mode of action. J Anim Sci 81(14):120–132. https://doi.org/10.2527/2003.8114_suppl_2E120x

Li F, Neves ALA, Ghoshal B, Guan LL (2018) Symposium review: mining metagenomic and metatranscriptomic data for clues about microbial metabolic functions in ruminants. J Dairy Sci 101(6):5605–5618. https://doi.org/10.3168/jds.201713356

Leahy SC, Kelly WJ, Ronimus RS, Wedlock N, Altermann E, Attwood GT (2013) Genome sequencing of rumen bacteria and archaea and its application to methane mitigation strategies. Animal 7(Suppl. 2):235–243. https://doi.org/10.1017/S1751731113000700

Lassen J, Løvendahl P (2016) Heritability estimates for enteric methane emissions from Holstein cattle measured using noninvasive methods. J Dairy Sci 99(3):1959–1967. https://doi.org/10.3168/jds.2015-10012

Lassen J, Poulsen NA, Larsen MK, Buitenhuis AJ (2016) Genetic and genomic relationship between methane production measured in breath and fatty acid content in milk samples from Danish Holsteins. Anim Prod Sci 56(3):298–303. https://doi.org/10.1071/AN15489

Martin C, Morgavi DP, Doreau M (2010) Methane mitigation in ruminants: from microbe to the farm scale. Animal 4(3):351–365. https://doi.org/10.1017/S1751731109990620

McAllister TA, Newbold CJ (2008) Redirecting rumen fermentation to reduce methanogenesis. Aust J Exp Agric 48(1–2):7–13. https://doi.org/10.1071/EA07218

McAllister TA, Beauchemin KA, Alazzeh AY, Baah J, Teather RM, Stanford K (2011) Review: the use of direct fed microbials to mitigate pathogens and enhance production in cattle. Canadian J Ani Sci 91(2):193–211. https://doi.org/10.4141/cjas10047

Moss AR, Jouany J, Newbold J (2000) Methane production by ruminants: its contribution to global warming. Ann Zootech 49(3):231–253. https://doi.org/10.1051/animres:2000119

Morgavi DP, Kelly WJ, Janssen PH, Attwood GT (2013) Rumen microbial (meta)genomics and its application to ruminant production. Animal 7Suppl1:184–201. https://doi.org/10.1017/S1751731112000419

Newbold CJ, de la Fuente G, Belanche A, Ramos-Morales E, McEwan N (2015) The role of ciliate protozoa in the rumen. Front Microbiol 6:1313. https://doi.org/10.3389/fmicb.2015.01313

OECD/FAO (2019) OECD-FAO Agricultural Outlook 2019–2028. OECD Publishing, Paris. https://www.oecd-ilibrary.org/agriculture-and-food/oecd-fao-agricultural-outlook-2019-2028_agr_outlook-2019-en. Accessed 15 Mar 2022. https://doi.org/10.1787/agr_outlook-2019-en

Patra A, Park T, Kim M, Yu Z (2017) Rumen methanogens andmitigation of methane emission by anti-methanogenic compounds and substances. J Anim Sci Biotechnol 8:1–13. https://doi.org/10.1186/s40104-017-0145-9

Rojas-Downing MM, Nejadhashemi AP, Harrigan T, Woznicki SA (2017) Climate change and livestock: impacts, adaptation, and mitigation. Clim Risk Manag 16:145–163. https://doi.org/10.1016/j.crm.2017.02.001

Re`nger A, Clark H, Cowie AL, Emmet-Booth J, Gonzalez Fischer C, Herrero M, Howden M, Leahy S (2021) How necessary and feasible are reductions of methane emissions from livestock to support stringent temperature goals? Phil Trans R Soc A 379:20200452. https://doi.org/10.1098/rsta.2020.0452

Roehe R, Dewhurst R, Duthie C-A, Rooke J, McKain N, Ross D, Hyslop JJ, Waterhouse A, Freeman TC, Watson M, Wallace RJ (2016) Bovine host genetic variation influences rumen microbial methane production with best selection criterion for low methane emitting and efficiently feed converting hosts based on metagenomic gene abundance. PLoS Genet 12(2):1–28. https://doi.org/10.1371/journal.pgen.1005846

Russell JB, Wallace RJ (1997) In: Hobson PN, Stewart CS (eds) Energy-yielding and energy-consuming reactions in the rumen microbial ecosystem. Blackie Academic and Professional, London, pp 246–282

Shahzad K, Loor JJ (2012) Application of top-down and bottom-up systems approaches in ruminant physiology and metabolism. Curr Genomics 13:379–394. https://doi.org/10.2174/138920212801619269

Suravajhala P, Kogelman LJ, Kadarmideen HN (2016) Multi-omic data integration and analysis using systems genomics approaches: methods and applications in animal production, health and welfare. Genet Sel Evol 48:38. https://doi.org/10.1186/s12711-016-0217-x

Tapio I, Snelling TJ, Strozzi F, Wallace RJ (2017) The ruminal microbiome associated with methane emissions from ruminant livestock. J Anim Sci Biotechnol 8:7. https://doi.org/10.1186/s40104-017-0141-0

Tamburini A (2020). Mitigation strategies of environmental impact of animal production: AGRIFOOD LCA lab. URL: https://sites.unimi.it/dottorato_aab/files/2020-06-25-Tamburini%2D%2D-Mitigation-Strategies-of-environmental-impact-of-animal-production.pdf. Accessed 15 Mar 2022

Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI (2007) The human microbiome project. Nature 449(7164):804–810. https://doi.org/10.1038/nature06244

UNEP (2021). Methane emissions are driving climate change. Here's how to reduce them. https://www.unep.org/news-and-stories/story/methane-emissions-are-driving-climate-change-heres-how-reduce-them. Accessed 15 Mar 2022

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63. https://doi.org/10.1038/nrg2484

Wang H, Zheng H, Browne F, Roehe R, Dewhurst RJ, Engel F, Hemmje M, Lu X, Walsh P (2017) Integrated metagenomic analysis of the rumen microbiome of cattle reveals key biological mechanisms associated with methane traits. Methods 124:108–119. https://doi.org/10.1016/j.ymeth.2017.05.029

Wang K, Nan X, Chu K, Tong J, Yang L, Zheng S et al (2018) Shifts of hydrogen metabolism from methanogenesis to propionate production in response to replacement of forage fiber with non-forage fiber sources in diets in vitro. Front Microbiol 2764(9):2764. https://doi.org/10.3389/fmicb.2018.02764

Weimer PJ, Stevenson DM, Mantovani HC, Man SLC (2010) Host specificity of the ruminal bacterial community in the dairy cow following near-total exchange of ruminal contents. J Dairy Sci 93(12):5902–5912. https://doi.org/10.3168/jds.2010-3500

Yan T, Mayne CS, Gordon FG, Porter MG, Agnew RE, Patterson DC, Ferris CP, Kilpatrick DJ (2010) Mitigation of enteric methane emissions through improving efficiency of energy utilization and productivity in lactating dairy cows. J Dairy Sci 93(6):2630–2638. https://doi.org/10.3168/jds.2009-2929

Chapter 14
Harnessing Knowledge from COVID-19 Scenario for New Generation Vaccine Development to Control Pandemics in Animals

Ablesh Gautam, Himani Gautam, Ankita Sharma, and Yashpal Singh Malik

Abstract Vaccines are the most efficient and economic tool to control and prevent infectious diseases in animals. First-generation vaccines; live attenuated and inactivated vaccines have been successful in producing healthy and productive livestock and controlling various infectious diseases. However, with increasing rate of emerging infections and the risk of resurgence of vaccine-controlled diseases, new vaccine strategies that are safer to use, and implicit both humoral and cellular immune responses need to be devised. Older generation vaccines are not capable of controlling the urgencies caused by newer epidemics and pandemics. The next-generation vaccine platforms such as recombinant subunit, virus-like particles, DNA, RNA, viral-vectored vaccines, etc., are capable of eliciting both humoral and cellular immune response, are cost-effective, and safer to use. With the emergence of COVID-19 pandemic, the field of medical science has excelled in producing fastest vaccines ever using next-generation vaccine technologies and reverse vaccinology. Similarly, the information of new generation vaccine platforms can be exploited in devising better vaccine development strategies in the field of veterinary science for controlling any such pandemics among animals in near future. This chapter elaborates upon various old generation and new generation vaccine platform technologies, reverse vaccinology, their role in veterinary vaccinology, and how these newer vaccine platforms can be deployed to design novel, efficient, and safer vaccines to control pandemics in animals, thereby also preventing zoonotic diseases under one-health program.

A. Gautam (✉) · A. Sharma
Central Research Institute, Kasauli, Himachal Pradesh, India

H. Gautam · Y. S. Malik
College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India
e-mail: malikyps@gadvasu.in

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_14

Keywords COVID-19 · Vaccine · Next-generation vaccines · Pandemics in animals · Reverse vaccinology

14.1 Introduction

Vaccination represents the most cost-effective method for the prevention and control of livestock diseases and the most widely used tool in veterinary medicine. More recently, with the world observing the COVID-19 pandemic, the development of vaccines has provided a ray of hope by providing protective immunities to the masses. With the vaccines developed in the past, various livestock diseases have been eradicated such as rinderpest. Veterinary vaccines are important not only in controlling livestock diseases, but also prevent transmission of zoonotic and foodborne infections. Table 14.1 depicts various vaccine technologies currently employed in veterinary medicine for various animal species. Conventional approaches to vaccine development have ruled over the market for many years; however, the nonconventional next-generation vaccines is the ideal approach in combatting pandemics in animals as well as COVID-19 in humans.

14.2 History of Vaccine Development

History of vaccination dates back to 1000 BC in China, where the smallpox endemic was usually treated by people through variolation, i.e., inhalation of dried crusts from smallpox lesions or inoculation of pus from a lesion into a scratch on the forearm of young children (Leung 2011). In the late eighteenth century, Edward Jenner observed and studied Miss Sarah Nelmes, a milkmaid who had previously caught cowpox and was subsequently found to be immune to smallpox. This observation led him to do an experiment. On 14th May 1976, Edward Jenner used cowpox-infected material obtained from the hand of a milkmaid to successfully vaccinate 8-year-old James Phipps. Later, on first July, Jenner challenged the boy by deliberately inoculating him with material from a real case of smallpox and observed that he was not infected. Jenner developed the first vaccine based on these findings (Riedel 2005). Since then, vaccination has been used to overcome various emerging, re-emerging, and stable infectious diseases. Vaccination must confer humoral and/or cell-mediated immunity to produce immunological memory and confer protection against subsequent infections. The history of vaccinology can be divided into two phases based on developments: the classical first-generation vaccines reflect trials-and-error approaches that relied on the ability to isolate, attenuate, inactivate, and fractionate the pathogen of interest to generate antigens for a vaccine (De Gregorio and Rappuoli 2014). These inactivated (IV) and modified-live vaccines (MLV) have proved to be of greatest use for humans and animals against their pathogens for many years. The inactivated vaccines are safe and relatively inexpensive to produce, but predominantly present antigens via the MHC-II pathway and induce humoral

Table 14.1 Various vaccine technologies in veterinary medicine

Canine			
Technology	Pathogen	Vaccine	Manufacturer
Canine			
Subunit	Leishmania	CaniLeish®	Virbac
Recombinant protein	Leishmania	Leish-Tech®	CEVA Animal Health
Chimeric protein	Leishmania	Lentifend®	Laboratories Leti
Chimeric protein	Borrelia burgdoferi	Vangaurd® crLyme	Zoetis
Viral-vector(canary-pox)	Canine distemper virus	Recombiteck® CDV	Boehringer Ingelhim
DIVA vaccines	Leishmania	Leish-Tech®	CEVA Animal health
Swine			
Subunit	A. Pleuropneumoniae	Porcilis®APP	Merck Animal Health
Subunit	A. Pleuropneumoniae	Pleurostar APP	Novartis
Subunit (BEVS)	Porcine circovirus type 2	Pleurostar APP	Boehringer Ingelheim
Subunit (BEVS)	Porcine circovirus type 2	Circumvent® PCV-M G2	Merck Animal Health
Subunit (BEVS)	Porcine circovirus type 2	Procilis® PCV	Merck Animal Health
Subunit (BEVS)	Porcine circovirus type 2	CircoGard	Pharmgate Animal Health
Subunit (BEVS)	Classical swine fever	PorcilisPesti®	Merck
Subunit (BEVS)	Classical swine fever	Bayovac CSF E2®	Bayer
Chimeric viral-vector (PCV)	Porcine circovirus type 2	FosteraTMPCV	Zoetis
Chimeric viral-vector (BVDV)	Classical swine fever virus	Suvaxyn® CSF Marker	Zoetis
RNA replicon (VEEV)	Porcine circovirus type 2	iPED+	Merck Animal Health
RNA replicon (VEEV)	Swine influenza A virus	iPED+	Merck Animal Health
DIVA vaccines	Suid herpesvirus 1	Porcilis® begonia	Merck Animal Health
DIVA vaccines	Suid herpesvirus 1	Auskipra® GN	Hipra
Avian			
Viral-vector (fowl pox)	Avian influenza	Trovac®-AIV H5	Boehringer Ingelheim
Chimeric viral-vector (HVT/MD)	Avian influenza	Vectormune® AI	CEVA Biomune
Chimeric viral-vector (HVT/MD)	Newcastle disease	Vectormune® ND	CEVA Biomune

(continued)

Table 14.1 (continued)

Canine			
Technology	Pathogen	Vaccine	Manufacturer
Chimeric viral-vector (fowl pox)	Infectious Laryngotracheitis virus	Vectormune® FP LT	CEVA Biomune
Chimeric viral-vector (fowl pox)	Mycoplasma gallisepticum	Vectormune® FP MG	CEVA Biomune
Chimeric viral-vector (fowl pox)	Newcastle disease	Vectormune® FP-N	CEVA Biomune
Chimeric viral-vector (HVT/MD)	Newcastle disease	Innovax®-ND	Merck Animal Health
Chimeric viral-vector (HVT/MD)	Newcastle disease and infectious bursal disease	Innovax®-ND-IBD	Merck Animal Health
Chimeric viral-vector (HVT/MD)	Newcastle disease and infectious Laryngotracheitis	Innovax®-ND-ILT	Merck Animal HealthAdt.A24 FMD
Bovine			
Viral-vector (adenovirus)	Foot and mouth disease	Adt.A24 FMD	
DIVA	Foot and mouth disease	Adt.A24 FMD	GenVec
DIVA	Bovine Herpesvirus-1	Bovilis® IBR Marker Live	Intervet
DIVA	Bovine Herpesvirus-1	Hiprabovis® IBR Marker Live	Hipra
DIVA	Bovine Herpesvirus-1	Bayovac IBR Marker Vivum	Bayer
DIVA	Bovine Herpesvirus-1	Bayovac IBR Marker Inactivum	Bayer
DIVA	Bovine Herpesvirus-1	Rispoval® IBR-Marker Inactivated	Zoetis
DIVA	Bovine Herpesvirus-1	Rispoval® IBR-Marker Live	Zoetis
Feline vaccine			
Viral-vector (carnarypox)	Feline leukemia virus	PureVAX® Recombinant FeLV	Boehringer Ingelheim
Viral-vector (carnarypox)	Rabies	PureVAX® Feline Rabies	Boehringer Ingelheim
DIVA	Feline leukemia virus	PureVAX® Recombinant FeLV	Boehringer Ingelheim
Equine			
Viral-vector (canarypox)	Equine influenza	ProteqFlu	Boehringer Ingelheim
Viral-vector (canarypox)	West Nile virus	ALVAC®-WNV	Pfizer

(continued)

Table 14.1 (continued)

Canine			
Technology	Pathogen	Vaccine	Manufacturer
Rabbits			
Chimeric viral-vector (myxoma virus)	Rabbit hemorrhagic disease	Novibac® Myxo-RHD	Merck Animal Health
Chimeric viral-vector (myxoma virus)	Rabbit hemorrhagic disease	Novibac® Myxo-RHD plus	Merck Animal Health
Wildlife			
Viral-vector (human adenovirus type 5)	Rabies	ORNAB®	Artemis Technologies, Inc.,
Viral-vector (vaccinia virus)	Rabies	Raboral V-RG®	Boehringer Ingelheim

immunity, due to which strong pathogens may escape the pressure elicited by the vaccine (Melnick 1978). The modified live vaccines overcome this issue by successfully replicating in the host and eliciting both MHC-I and MHC-II pathways. These modified/attenuated pathogens mimic the natural infection. However, they pose a risk to animals by regaining pathogenicity (Meeusen et al. 2007).

With the advent of recombinant-DNA technology, second- and third-generation vaccines have shown great success, thereby paving the way to advancement in medicine. Subunit elements, conjugated/recombinant antigens, and synthetic proteins make up the second-generation vaccines. Recombinant subunit vaccines use overexpression for antigen production rather than using the virus itself. These subunit vaccines use full-length major antigenic proteins or a protein fragment or combination of protein fragments with a carrier protein as the antigen for inducing a strong immune response. However, these lack the pathogenic molecular patterns that are used by the immune system to recognize and memorize pathogens. These vaccines require adjuvants in the formulation and repeated administration to establish immune responses. Further, they are generally recognized by antigen-presenting cells (APCs) and presented via MHC-II complexes and, therefore, trigger humoral immune responses.

The gene-based vaccines, viral-vector vaccines, and live or inactivated chimeric vaccines constitute the third-generation vaccines. DNA- and RNA-based vaccines make use of plasmids delivered through the injection. DNA vaccines utilize a plasmid containing the gene that employs the host cellular machinery to encode the antigen of interest and elicit the immune response (Ferraro et al. 2011). The injection needs to be accompanied by electroporation for efficient adsorption of plasmids into the cells. These vaccines induce both humoral and cell-mediated immunity, are safe to administer without risk of re-emergence, and can be used as a broad-spectrum combination vaccine by combining multiple plasmids. However, they are poorly immunogenic, requiring repeated administration and an adjuvant. There are a few risks associated with the formation of anti-DNA antibodies and promoting mutations in the host genome causing dysplasia (Chapman and Rybicki

2019). The mRNA vaccines are advantageous for their low cost and rapid manufacturing process. Nanotechnology is used for vaccine delivery, wherein the antigen encoding mRNA is coated with lipid nanoparticles (LPNs) as a stable lipid bilayer, which is then transferred into the host cells. By expressing viral antigens, they induce adaptive immunity to produce neutralizing antibodies against the target antigen. The adverse effects of RNA vaccines include induction of autoimmune disorders through the development of type I IFNs and an extreme immune reaction of the nonhuman polynucleotide. We now find ourselves in an era of novel recombinant viral-vector vaccine technology that is genetically engineered and involves the insertion of DNA encoding key antigens into a viral vector. These vaccines are bioengineered viral-vectored vaccines that use the reproduction ability to express and clone the pathogenic antigens. The common viral vectors that are used are lentivirus, retroviruses, adenoviruses, and adeno-associated viruses. They are malleable, safe, and capable of producing strong cellular responses without the need for an adjuvant. Chimeric recombinant vector vaccines are another type of vaccines. The foreign genes of interest from multiple types of pathogens are inserted into self-replicating RNA replicons. These replicons get engulfed by dendritic cells upon inoculation. The antigens, thus after translation, get presented to the immune cells.

In veterinary medicine, the DIVA technology (Differentiation of Infected and Vaccinated Animals) is a critical tool in disease control and eradication. It is based on recombinant deletion of mutants of wild-type pathogens, wherein gene segments expressing viral proteins have been removed. The DIVA vaccines have been of great success in the control of foot and mouth disease, classical swine fever, bovine rhinotracheitis, etc. This has proved to be a safe, rapid, sensitive, and inexpensive method for determining pathogen-free flocks and herds (Suarez 2012).

14.3 Conventional Veterinary Vaccines

14.3.1 Live Attenuated Veterinary Vaccines

Live attenuated vaccines are produced by multiple passages of the pathogens (virus or bacteria) in unnatural hosts or cell lines. This process of passage of the infectious viral or bacterial strain renders the pathogen undergo random mutations, thereby losing its virulence while retaining its immunogenicity (Meeusen et al. 2007). The pathogen, however, retains its replicative behavior. The live attenuated vaccines are advantageous over other traditional vaccines in terms of their ability to induce both cellular as well as humoral immunity, and commonly do not require an adjuvant. Thus, the immunity attained lasts for a lifetime. However, these vaccines have multiple drawbacks. More often, the live strains used in the vaccines are not highly protective. Moreover, there is always a fair chance of reverting these attenuated strains into even more virulent phenotypes. Further, since, these vaccines carry live strains, the vaccines need refrigerated storage. Other disadvantages include local or other unwanted reactions after vaccine inoculation, challenges faced in effectively

culturing the virus or bacteria, and the possibility of eliciting an autoimmune response (Babiuk et al. 2003).

The live attenuated viral vaccines are challenging to produce. The complex process of production involves propagating the viral strains in living cells and the cell lines from different origins. Achieving viral titer and standardization in the living cells also pose challenges. The nature of macromolecular complexity of bacteria and viruses, for example, enveloped or non-enveloped viruses, further exhibits complexities in formulation of the vaccine. Since, live attenuated vaccines do not require adjuvants in general, they need less downstream processing in formulation in comparison to inactivated vaccines (van Gelder and Makoschey 2012).

14.3.2 Inactivated Veterinary Vaccines

Inactivated vaccines are made up of inactivated bacterins and/or toxins of one or more bacteria or serotypes, or killed viral strains (Meeusen et al. 2007). These vaccines are formulated in oil or aluminum hydroxide adjuvants. Inactivated vaccines are relatively less expensive to produce as compared to their counterpart live vaccines. Since these vaccines are stable, they are easy to handle under field conditions. These vaccines are inactivated using physical or chemical treatments, resulting in damage to the proteins, antigens, and nucleic acid of the vaccine pathogen. On account of this, the inactivated vaccines, though qualify for better safety profiles, nevertheless, do not provide effective long-term protection as seen in live vaccines (Cho et al. 2002).

For viral vaccines, the viruses are grown in live host cells, i.e., cell cultures, either in large roller bottles or bioreactors. The viral strains are then inactivated by physical or chemical treatment. Such types of treatments result in either inactivation of proteins and/or destruction of the nucleic acids. The inactivated antigen, thus obtained, is then purified, and formulated with optimum levels of adjuvant (Fig. 14.1).

14.3.3 Subunit Veterinary Vaccines

Subunit vaccines are prepared using either of the immunogenic parts of the vaccine pathogen. The immune response thus is elicited against that specific component only (Fig. 14.2). Polysaccharide subunit vaccines are comprised of long-chained carbohydrate molecules from the surface capsule of the bacteria. These vaccines are deficient in additional antigenic parts of the vaccine pathogen that induce T cell stimulation. The polysaccharides used in subunit vaccines are incapable of inducing T-cell stimulation sufficiently. This problem has been addressed by improvising these subunit vaccines using biotechnological applications. The polysaccharide

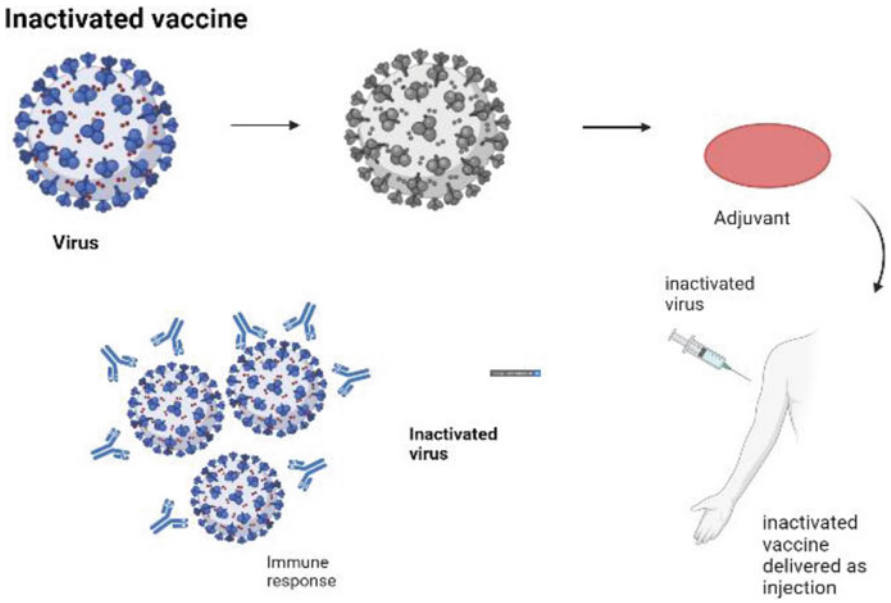

Fig. 14.1 Inactivated vaccine platform. Physically or chemically damaged antigens and nucleic acids formulated in adjuvants are delivered via injection

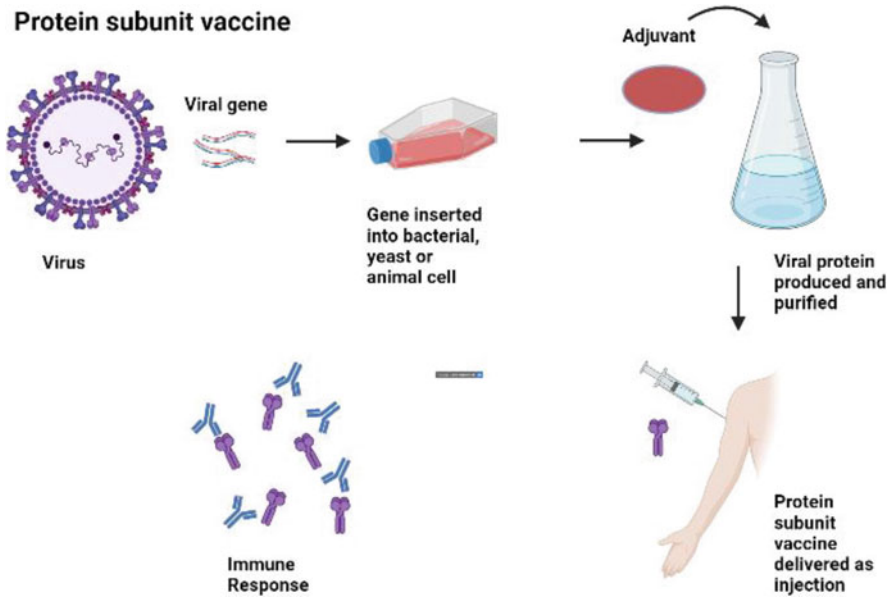

Fig. 14.2 Protein subunit vaccine platform. Viral genes inserted in bacterial or other host cells encode viral proteins that are delivered via injection

antigen is covalently linked to a carrier protein, often a toxoid, such as diphtheria or tetanus toxoid, resulting in polysaccharide–protein–conjugate vaccines. These conjugate vaccines have proved to elicit better immune responses by recruiting sufficient T cells (Dintzis 1992).

14.3.4 Toxoids

Often, pathogenicity is caused by toxins released by bacteria, rather than the pathogen alone. In such cases, vaccination has proved to be the most effective measure to combat such diseases caused by bacterial toxins. The native bacterial toxins are inactivated (toxoids) using chemical methods, purified and formulated with commercial adjuvants. These vaccines prepared using toxoids are efficient and have been used commercially. If produced conventionally, the production process has limitations in terms of the amount of toxin production in vitro. While some toxins produced can be highly toxic and, therefore, require high levels of biosafety procedures (Arimitsu et al. 2004).

The use of recombinant technology has overcome these limitations. With the technology, these toxins can be produced efficiently at an industrial scale while decreasing their reactogenicity and toxicity. For instance, the recombinant *Escherichia coli* toxins can be produced and inactivated with formaldehyde in merely 2–3 days. Since the toxic domain of the protein can be removed using recombinant technology, this production process does not require high levels of biosafety (Moreira et al. 2016).

14.4 New Generation Vaccines

14.4.1 Recombinant Subunit Veterinary Vaccines

Subunit vaccines are developed using a short part of antigenic sites or proteins that are noninfectious, nevertheless, capable of inducing a potent immune response. Such antigens used in vaccines are not capable of undergoing replication inside the host and, therefore, are advantageous over live vaccines. These vaccines offer enhanced safety as the whole pathogen need not be handled or administered. The production levels can be easily scaled up as it is easier to purify the expressed protein in larger quantities.

For the production of recombinant subunit vaccines, the gene encoding the antigen of interest is amplified from the genomic DNA of the pathogen and is cloned for expression and industrial-scale production. For heterologous protein expression, *Escherichia coli* is the widely used expression vector. This expression system, however, is not capable of introducing posttranslational modifications to the protein such as folding, glycosylation, etc. (Heinson et al. 2015; Simionatto et al. 2010). If

posttranslational modifications are imminent in vaccine candidate proteins, *Pichia pastoris*, methylotrophic yeast, can be used as an expression system. The yeast expression system is inexpensive and works efficiently for the heterologous production of recombinant proteins. This expression system enables necessary genetic modification, expression and secretion of proteins of interest, and posttranslational modifications. With yeast as an expression host, the yield can be achieved on an industrial scale (Ghosh and Nagar 2014; Hartwig et al. 2010). The heterologous expression system bypasses the risks of handling virulent or partially virulent strains of the pathogen, each time during production activities. Once the expression construct is prepared, it can be stored under optimum storage conditions and can be propagated as and when required. Thus, the heterologous expression system offers safety to the manufacturer, vaccine handlers, as well as end-users. With the expression of proteins in live hosts, the native form of proteins such as their conformations, protein folding, conformational epitopes, posttranslational modifications, etc., is preserved in subunit vaccines (Eshghi et al. 2009).

Another mammalian expression system using insect cells to express recombinant proteins also allows posttranslational modifications such as glycosylation, phosphorylation, and signal peptide cleavage. This approach has successfully been used to develop commercial vaccines for porcine circovirus type 2 and classical swine fever (Aida et al. 2021; Crisci et al. 2012).

Furthermore, expression systems using plants or plant cell lines have also been used for the production of veterinary and human vaccines. These systems are advantageous for inexpensive large-scale production of recombinant proteins in a relatively shorter period. This expression system was fruitfully used in tobacco plants for the first time, resulting in the first commercial vaccine against New Castle Disease Virus (NDV) in poultry. This system has further been explored for other diseases of livestock as well, such as Infectious Bronchitis Virus (IBV), Infectious Bursal Disease Virus (IBDV) in poultry, Enterotoxigenic *Escherichia coli* (ETEC) in farm animals, Bovine Viral Disease (BVD), and Bovine Herpes Virus in cattle (Francis 2018).

More than one protein can be incorporated in recombinant subunit vaccines, thereby enabling a broader immune response against similar strains or serotypes of pathogens (Odir et al. 2011). Amid the advantages of subunit vaccines, there are limitations as well. Since only a part of the protein is used in subunit vaccines, the immune response generated does not equate to the immune response generated by using live or inactivated whole organisms. Potent adjuvants and booster doses are required in vaccine formulations to enhance the immune response in the animals (Plotkin 2014). Other approaches that are used to increase immunogenicity and stability of the subunit vaccines include virus-like particle (VLP) vaccines and nanoparticle (NP) vaccines.

The virus-like particles (VLPs) are spontaneously assembled particles formed from several viral structural proteins that act as vaccine delivery agents. Compared to attenuated viruses, VLPs are safer, more effective, have higher immunogenicity, excellent adjuvant characteristics, and can induce both humoral and cellular immune responses.

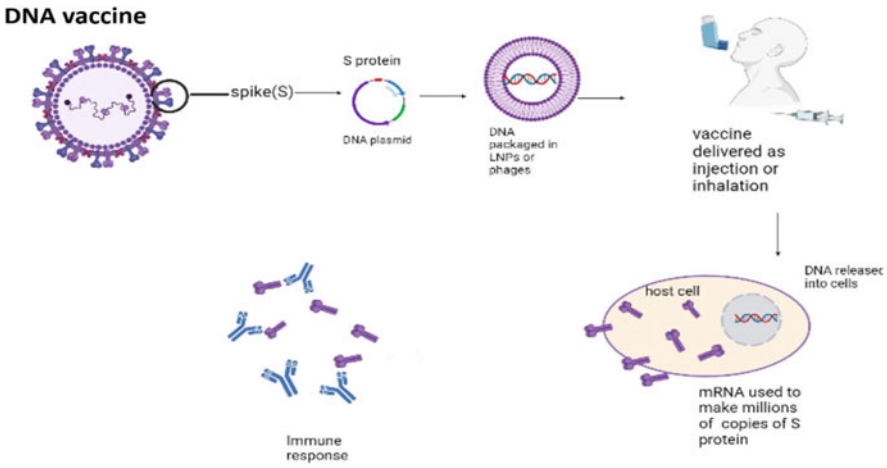

Fig. 14.3 VLPs vaccine platforms. Viral proteins expressed in plant or cell system form self-assembly of VLPs that are delivered along with adjuant via injection

The VLP vaccines are based on using a mechanism that enables the production of certain particles that has the self-assembling abilities of true viruses (Fig. 14.3). These vaccines are produced by expressing vaccine candidate antigens into a prokaryotic or eukaryotic expression system that results in the generation of particles that along with the inherent properties of the antigens, can, like viruses, assemble on their own (Kushnir et al. 2012). In another method for the production of VLP vaccines, the VLP template can be made and chemically linked to the antigens onto the premade particles (Kushnir et al. 2012). Since these VLP vaccines lack actual viral genomes, they are deficient in replicating inside the host cells. Therefore, these vaccines offer a better safety profile as compared to live viral vaccines (Mohsen et al. 2017). Nevertheless, the VLP vaccines are capable of inducing cellular and humoral immune responses in the vaccinated population (Chackeriam 2007; Grgacic and Anderson 2006). The dendritic cells take up the VLPs and present them onto MHC class I and II to induce an adaptive immune response. Later, $CD8^+$ T cells and $CD4^+$ helper cells are activated resulting in cell-mediated immune response and B cell activation (Deml et al. 2005; Kushnir et al. 2012; Murata et al. 2003; Paliard et al. 2000; Schirmbeck et al. 1996; Win et al. 2011). In humans, the VLP vaccines are now studied for numerous viral diseases such as hepatitis B vaccine (HBV) Engerix (GlaxoSmithKline 1989), the human papillomavirus vaccine (HPV) Cervarix (GlaxoSmithKline 2009) the HBV vaccine Recombivax (CDC 2018), the HPV vaccine Gardasil (Merck&Co. 2015), etc. The VLP vaccines for malaria (Agnandji et al. 2012; Heinson et al. 2015), influenza (Song et al. 2011), rotavirus (El-Attar et al. 2009; Zhou et al. 2011), tuberculosis (Dhanasooraj et al. 2016), Zika virus (Boigard et al. 2017), and HIV (Franco et al. 2011; Pillay et al. 2010) are in clinical trials.

Nanoparticle vaccines (NPs) are produced by binding expressed and purified antigen with a carrier molecule, thereby enhancing its immunogenicity and stability (Frietze et al. 2016). These carrier molecules can be categorized as organic molecules, which are mostly lipid-based, and inorganic molecules, which are mostly polymer- or metal-based (Han et al. 2018; Li et al. 2017a, b; Pati et al. 2018; Zhao et al. 2014). The NPs are more stable than VLPs; however, the innate immune response achieved by VLPs is better than that achieved by NPs. The immunogenicity of NPs can be increased to some extent by customizing the carrier molecule in terms of its size, surface charge, shape, and hydrophobicity (Al-Halifa et al. 2019; Pati et al. 2018; Zhao et al. 2014). The carrier molecules may also be designed to improve antigen presentation by antigen-presenting cells (APCs) (Dhanwani et al. 2017; Zhao et al. 2014).

Vaccine development using this technology is being investigated for various livestock diseases such as Bovine Viral Diarrhea Virus (BVDV-1), Bovine Respiratory Syncytial Virus (BRSV), Bovine Para Influenza 3 Virus (BPI3V), BVH-1, *Mycobacterium avium* subspecies paratuberculosis, *Brucella abortus,* and *Anaplasma marginale* (Maina et al. 2020).

14.4.2 Vectored Veterinary Vaccines

For conventional vaccines, the antigens were delivered into the host body either along with live or inactivated microorganisms or in the form of inactivated toxins. With newer technology, advancements in antigen/gene delivery systems, and recombinant technology, vaccine development has reached the next level. A class of novel vaccines is prepared by using vectors to deliver the vaccine candidate genes/antigens into the host system. This technology is referred to as vector vaccine technology. Such vector itself serves as immunogenic and can deliver multiple antigens into the host. These vaccines do not require an adjuvant to elicit a good immune response (Fig. 14.4).

The vectored vaccines that are prepared using recombinant technology can be categorized as live vector vaccines or naked DNA vaccines. For recombinant vector vaccine production, DNA encoding vaccine candidate antigen from the pathogen is inserted into the vector plasmid using molecular cloning techniques. The recombinant plasmid, thus obtained, is then transfected into the live vector, which is then grown in culture media or cell culture cells. This whole process gives rise to recombinant vector vaccines (Tiwari et al. 2020).

Live vectored vaccines can be developed using live attenuated bacteria or viruses that can display potent immunogenic antigens or other microorganisms. These live vectors are beneficial in generating their natural immunity as well. Presently, poxviruses are the most commonly used recombinant vector. The large genome size of these viruses vary from 130 to 300 kbp (Panicali and Paoletti 1982; Tiwari et al. 2020), enabling larger foreign DNA sequences inserted into their genomes. Numerous vectored veterinary vaccines have been developed by using vaccinia

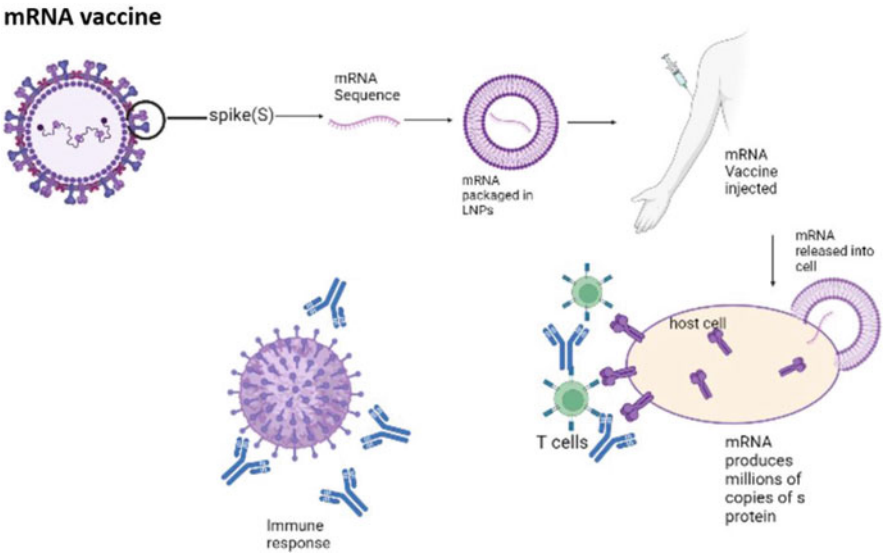

Fig. 14.4 Viral-vector vaccine platform. The vectored vaccines prepared using recombinant technology delivers the plasmid containing DNA encoding for pathogenic antigen into the host cell

virus, fowl pox virus, capri poxvirus, parapoxviral, and canarypox viruses as live vectors (Tiwari et al. 2020). The canarypox virus vector system has served to develop veterinary vaccines against West Nile virus, canine distemper, feline leukemia, rabies, and equine influenza (Jorge and Dellagostin 2017).

The attenuated version of the vaccinia virus, Modified Vaccinia Ankara (MVA), is a well-characterized strain that has proved to be a safer vaccine vector by losing its replicating behavior in mammalian cells (Volz and Sutter 2017). In experimental stages, the MVA bivalent vaccine has proved to be protective against Rift Valley Fever (RVF) and Bluetongue virus (BTV) in sheep (Calvo-Pinilla et al. 2020).

A variety of human and animal viruses has also been investigated to be used as vaccine vectors. The most recent and widely used now are the chimpanzee adenovirus ChAdOx1 and human adenovirus serotype 26 (Ad26) vectors that have been used to develop a human vaccine against SARS-CoV-2 (Mendonça et al. 2021).

Adenovirus has most of the qualities required for a successful vaccine delivery tool; they elicit strong humoral and cell-mediated antigen-specific responses. They are manipulable and allow alteration of replicative capacity, cell tropism, immunogenicity, and insertion of antigens of interest; although given the need for veterinary vaccines to be delivered to remote locations and stored under optimal conditions, substantial work is being done to prepare thermostable vaccines that can remain stable for up to 6 months at 45 °C (Chapman and Rybicki 2019).

The bacterial vector system has been more commonly used in recombinant BCG, which can deliver multiple antigens and induce potent immunity (Rizzi et al. 2012). The use of plants as vectors for vaccine development has prodigious potential in the

field of veterinary vaccinology. The vaccines thus produced by using transgenic plants can deliver antigens through animal feed (Shams 2005).

14.4.3 DNA and RNA Vaccines

The DNA and RNA vaccines are developed using the nucleic acids DNA or RNA.

These vaccines are prepared using molecular techniques and recombinant technology. A plasmid is constructed by inserting a gene of interest along with all the necessary elements for transcription and translation such as promoter, polyadenylation signal sequence, bacterial origin of replication, etc. This plasmid-based vaccine is transfected into the host cells (Fig. 14.5). The host cells after taking up the plasmid vaccine, start transcribing the gene into mRNA. Later, the host cells start expressing the antigens after undergoing translation. These antigenic proteins, thus produced, are recognized by the host immune system as foreign proteins and evoke cellular and humoral immune responses (Jorge and Dellagostin 2017).

These vaccines are economical to produce, are simpler in designing, and since, the whole organism need not be handled, these have high biosafety profile. The DNA vaccines exhibit enhanced stability as well. There are a few DNA vaccines used in veterinary medicine (Hobernik and Bros 2018). DNA vaccines are incapable of inducing potent immunity in humans as well as large animals (Khan 2013).

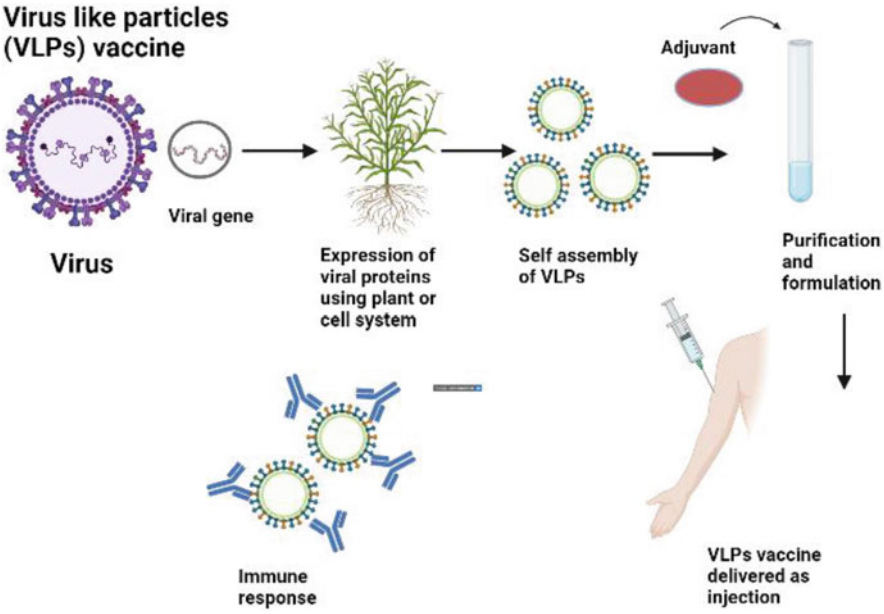

Fig. 14.5 DNA vaccine platform. Here plasmids are packaged in LNPs or phages and delivered as injection or inhalation followed by electroporation system

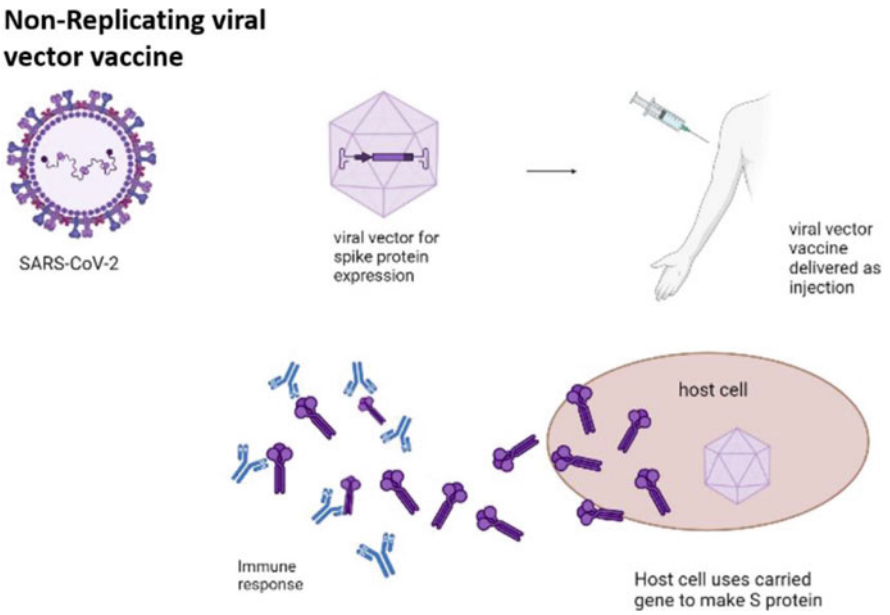

Fig. 14.6 mRNA vaccine platform the mRNA sequence encoding for spike proteins, packaged in LNPs are injected into host that produces millions of copies of a protein evoking cellular and humoral response

Self-amplifying (sa) RNA based on genetically engineered replicons that can be delivered as viral replicon particles (VRPs) is another related platform (Fig. 14.6). The saRNA produced after in-vitro transcription can express higher levels of antigens at lower doses, thus providing potential cost benefits. The Venezuelan equine encephalitis virus-derived alphavirus-based saRNA vector has been evaluated for veterinary vaccines for porcine endemic diarrhea and swine influenza (Entrican and Francis 2022).

In veterinary medicine, these nucleic acid vaccines are advantageous over other types of vaccines, especially for viral diseases. Preexisting immunity is of major concern in the case of vectored vaccines, while safety concerns are crucial in the case of live vaccines. Further, by exhibiting intracellular expression of antigens, the vaccines induce cytotoxic T cells (Meeusen et al. 2007).

14.4.4 Covid-19 Vaccine Development Using New Generation and Reverse Vaccinology

The spread of SARS-CoV-2 worldwide has given rise to alarming health conditions. Vaccinating people against the virus came out as an effective way of disease control. The prerequisites of the vaccine development process include information regarding

pathways that the pathogen exploits to invade the host cells and evade host immunity, candidate antigens that can be targeted for developing vaccines, immunization mechanisms, animal models, production processes, ability to be produced at industrial scale, estimation of the outbreak, target population, etc. (Wang et al. 2020). The process of vaccine manufacturing is not limited to only these factors. After a successful review of the candidate antigens, research and thorough study on choosing the right type of vector or carrier, vaccine formulation, adjuvant, route of administration, dosage, storage conditions, etc., are other challenges that are faced during the development of a vaccine (Wang et al. 2020). In the case of SARS-CoV-2, the process of vaccine development was supported by the amalgamation of artificial intelligence, computational tools, and biotechnology (Chauhan et al. 2020; Sun et al. 2020; Zhou et al. 2020). The genomic matches to other coronaviruses further helped to select the most potent vaccine candidate antigens for developing the COVID-19 vaccine (Chauhan et al. 2020). Figure 14.2 depicts the various vaccine platforms in SARS-CoV-2 vaccine development.

For antigen selection, the S1/S2 protein subunits, receptor binding domain (RBD), and S protein/gene are the major target sites as proposed by researchers for vaccine development. Out of non-neutralizing antibodies which are generated to S, E, and M proteins, M and E proteins are considered to have poor immunogens for humoral responses (Dai and Gao 2021). Also, the enhanced humoral and T-cell mediated immune response could be targeted by the addition of structural (N) and/or nonstructural functional proteins as vaccine antigens (Dai and Gao 2021). The administration of vaccines in the host plays an important role in vaccine strategies. The most preferred routes are subcutaneous (SC), intradermal (ID), and intramuscular (IM) while common mucosal routes are noninvasive oral and nasal.

Different vaccine platforms are being used to develop vaccine candidates against COVID-19 such as traditional whole-pathogen vaccines like a live attenuated virus, inactive virus, protein subunit, virus-like particle, etc. On the other hand, next-generation vaccines include nucleic acid-based, recombinant viral vectors, antigen-presenting cells, and nanoparticle-based (Calina et al. 2020; Esmaeilzadeh and Elahi 2021; Jeyanathan et al. 2020). Figure 14.7 depicts the different vaccine platforms devised for SARS-CoV-2 vaccine production and Table 14.2 describes the different forms of COVID-19 vaccines along with their potential limitations.

14.4.4.1 Classical Vaccines

Live Attenuated

The two major components, antigens and infection signals, can be easily obtained from live attenuated vaccines to trigger and activate the host immune system (Jafari et al. 2022). Deletion or mutation of virulent genes can help in designing attenuated virus strain (Jafari et al. 2022). By using deletion of the structural E protein, many nonstructural protein vaccines have been engineered against several coronaviruses (Jeyanathan et al. 2020; Netland et al. 2010). Another method to produce attenuated

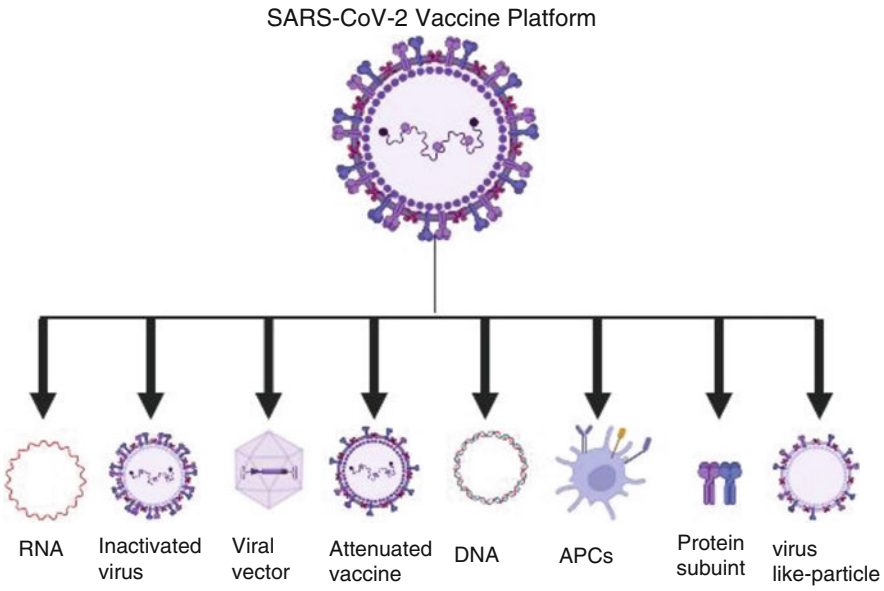

Fig. 14.7 Different vaccine platforms for development of SARS-CoV-2 vaccine production

strain is by using codon deoptimization. This method results in the slow translation of viral protein leading to a weak in vivo virus still capable of replicating in vitro, considering that the viral protein chosen for deoptimization is the right one (Jeyanathan et al. 2020). FDA-approved Calmette-Gu□rin (BCG) for tuberculosis is under investigation in clinical trials and is being repurposed to use against SARS-CoV-2 (Jafari et al. 2022).

Inactivated Viral Vaccine

In this method of vaccine approach, the inactivated virus is injected into the host to induce an immune response (Jafari et al. 2022). Thus, this can be a quick way to design vaccines by using well-established infrastructure and techniques in a pandemic situation like COVID-19 (Jeyanathan et al. 2020). The efficiency of such vaccines can be enhanced by using adjuvants and repeated administration. Moreover, there is a poor induction of $CD8^+$ T-cells by inactivated viral vaccines. This can be overcome by using TH1 cell-skewing modified alum or other adjuvants like CpG (Del Giudice et al. 2018). The vaccine BBIBP-CorV by Sinopharm (ChiCTR2000034780) is currently under phase-III trial for the SARS-CoV-2 virus (Jafari et al. 2022).

Table 14.2 Different forms of Covid vaccines along with their potential limitations

Covid vaccine platforms	CoVID vaccines	Potential limitations
Live attenuated virus	COVI-VAC (Codagenix/Serum Institute of India)	• Not highly protective • Reversion to virulence to a more virulent phenotype may occur • Refrigerated storage is required
Inactivated virus	Corona Vac(Sinovac); BBIBP-CorV; BBV152-Covaxin (Bharat Biotech)	• Cannot provide effective long-term protection • Unable to cope with the prevailing strains in the field • Poor induction of CD8 + T cells
Viral-vector (non-replicating)	ChAdOx1-S, Covishield, Vaxazervai (AsteraZeneca); Ad5-nCoV (Janseen pharmaceutical); Ad26.CoV2.S(Johnson and Johnson)	Continuous administration is required that leads to development of immune response Presence of pre-existing anti-vector immunity Episomal non-integration of vector genome into host genome Unintended transduction might cause cell toxicity and cell death
Viral vector (replicating) + APC	AV-COVID-19 (Avivita biomedical, Inc)	Potential for tumorigenesis
RNA-based vaccine	mRNA-1273(Moderna+ NIAID); BNT162b2(3 LNP-mRNAs)(Pfizer+ fosum pharma); CVnCoV (CureVac AG)	Risk of integration into host genome Unstable Challenges in adequate cellular uptake Induction of autoimmune disorders
DNA-based vaccine	nCov vaccine (Zydus Cadila); INO-4800 + electroporation (Inovio + Suzhou)	Incapable of inducing potent immunity Require repeated administration and adjuvants Risk of anti-DNA antibodies that promote mutations in host genome
Protein subunit	NVX-CoV2373 (Novavax); Razi Cov Pars (Razi Vaccine and Serum Research Institute)	Lack pathogenic molecular patterns Require adjuvants Repeated administration
Virus-like particles	SARS-CoV-2 VLP Vaccine: Vaccine Wuhan; Vaccine-Alpha; Vaccine-Wuhan + Alpha Variant (The Scientific and Technological Research Council of Turkey)	Redosing overtime is necessary to maintain immunity

Protein Subunit Vaccine

A full-length SARS-CoV-2 S protein or a protein fragment like RBD or RBD combined with a carrier protein is used in protein subunit vaccine as an antigen for inducing a strong immune response in the host (Jafari et al. 2022). The drawback of the protein subunit vaccine is its inability to generate cell-mediated immune response due to the poor immunogenicity of proteins and peptides. Thus, like inactivated viral vaccines, they also require adjuvant and repeated administration to establish immune memory (Diamond and Pierson 2020).

This approach to vaccine development is applied in assessing the immunogenicity of full-length recombinant SARS-CoV-2 spike nanoparticle vaccine (NVX-CoV2373) with or without Matrix-M as the adjuvant (Wang et al. 2020). This vaccine Novavax against coronavirus is under phase-III trial.

Virus-like Particles

Virus-like particles (VLPs) have emerged as successful vaccine delivery agents. They are capable to initiate the 3D structure of native proteins, thus triggering the immune system (Jafari et al. 2022). Devoid of infectious genetic material and functional proteins, VLPs are considered a good option for vaccine production (Jeyanathan et al. 2020). When talking of coronaviruses, the viral proteins (S, E, and M with or without N) are co-expressed in eukaryotic producing cells to form VLPs (Jafari et al. 2022). Similar to the parent virus, VLPs can bind and enter cells presenting ACE2 by the S protein on their surface (Naskalska et al. 2018). Several vaccines are in trials using VLPs as platforms.

14.4.4.2 Next-Generation Vaccines

RNA Vaccines

When compared to other vaccines, mRNA vaccines are low in cost, noninfectious, rapidly manufactured, and easy to administer. The introduction of nanotechnology to mRNA vaccines has boosted their delivery via the ID or IM route into the host (Jafari et al. 2022). A stable lipid bilayer of antigen-encoding mRNA which is coated with lipid nanoparticles (LNPs) can be effectively transferred in vivo into the host cells (Salian et al. 2021).

In case of COVID-19 pandemic, some biotech companies such as Pfizer, BioNTech, Moderna, and CureVac have developed advanced mRNA vaccines (Jafari et al. 2022). Among these, Moderna's mRNA-1273 vaccine encoding a perfusion-stabilized spike protein encapsulated in LNPs was the first one to enter phase-I clinical trials (NCT04283461) (Frederiksen et al. 2020). The phase-III trials (NCT04470427) have shown 94% efficiency. Adding to its clinical trial (NCT04380701), four Pfizer's nucleoside-modified mRNA candidates

(BNT162a1, BNT162b1, BNT162b2, and BNT162c2) have been designed to test its safety, efficiency, and immunogenicity. The study also revealed that BNT162b1 produces a stronger CD8-T cell response, which leads to the production of CD4-T cells and neutralizing antibodies when compared with Moderna's candidate vaccine (Sahin et al. 2020).

DNA Vaccines

DNA vaccines are made up by using an engineered plasmid that encodes for the pathogen-specific antigen and a carrier vector for introducing the plasmid into the host cell (Jafari et al. 2022). The plasmid further induces the expression of vaccine antigen post-transduction into the host. To fight the COVID-19 infection, Inovio pharmaceutical designed a phase-I clinical trial (NCT04336410) for studying the effect of a synthetic DNA vaccine expressing SARS-CoV-2 S protein (INO-4800). Another method using the CRISPR/Cas9 system to generate mouse models expressing hACE2 to study the transmission and pathogenesis of SARS-CoV-2 was used by Sun et al. for developing a tool to evaluate COVID-19 vaccines (Sun et al. 2020).

Antigen-Presenting Cells

Considering the importance of APCs in the immune response to a vaccine, APCs are developed artificially by using pathogen-associated antigens (Jafari et al. 2022). Artificial antigen-presenting cells (APCs) expressing the viral protein activate T cells and induce an immune response against the virus when induced in the host cells (Li, Shao, et al., 2017). Inactivated APCs expressing conserved structural and protease epitopes of SARS-CoV-2 are under phase-I clinical trial (NCT04299724). The safety of this vaccine is being evaluated in healthy and COVID-19 positive volunteers (Al-Kassmy et al. 2020).

Recombinant Viral-Vectored Vaccines

The generation of a strong immune response irrespective of the presence of an adjuvant, safety and requirement of a single dose has made the use of recombinant viral-vectored vaccines more popular for infectious diseases (van Riel and de Wit 2020). Vaccine processing includes the process of scaling up in a bioreactor for replacing the adherent cells, which is followed by downstream processes such as removal of impurities using ion-exchange chromatography, DNA digestion, centrifugation, etc.

Ad26-vectored COVID-19 vaccine against SARS-CoV-2 was found to be enormously safe in the nonhuman primate model and is under phase-I (NCT04568811) and phase-II (NCT04566770) trials to evaluate the immunogenicity of the doses

(Jafari et al. 2022). Also, Ad5-nCoV is under trial (NCT04916886) in different age groups.

To test the safety, efficacy, and immunogenicity of the nonreplicating ChAdOx1 nCoV-19 (AZD1222) vaccine, a product of Oxford University in collaboration with pharmaceutical company Astra-Zeneca has undergone phase-III trial (NCT04536051), and phase-IV trials had started in February 2021 (Jeyanathan et al. 2020). For emergency immunization among adults aged 18 and above, AZD1222 has been licensed by the UK Medicines and Healthcare products Regulatory Agency (MHRA) (Jafari et al. 2022).

Nanoparticles-Based Vaccines

The use of NPs in therapeutics has emerged rapidly in the last decade. The antiviral property of NPs like metals, small interfering RNA (siRNA), peptides, graphene oxide, and organic materials has been used for the treatment of COVID-19 (Hu et al. 2017; Muhammad et al. 2020; Ye et al. 2015). The delivery of therapeutic contents by NPs containing immune-regulating molecules and antioxidants (adenosine and alpha-tocopherol, respectively) to inflammation sites can reduce inflammation, cytokines reactions, and oxidative stress related to COVID-19 (Dormont et al. 2020; Merad and Martin 2020). As proposed by computational approaches, SARS-COV-2 blocking could be done by ACE-2-based peptide inhibitors. Thus, hypothesize that virus activation in the lungs could be prevented by NPs mimicking the binding domain of the virus used as inhaling therapeutics (Han and Král 2020).

Virus-like nanoparticles (VLNPs)-based vaccines are more efficient and nontoxic (Jafari et al. 2022). They can also accommodate additional protein expression of multiple antigens as carriers on their surfaces (Moon et al. 2012). By using the SARS-CoV-1 S protein, antibody production and inhibition of coronavirus function were observed in self-assembled peptide NPs (Pimentel et al. 2009). The VLNPs vaccine using MERS-CoV S protein NPs with an adjuvant combination of the matrix protein (M1) was found to be efficient in inhibiting the MERS-CoV replication in the lungs of mice (Jafari et al. 2022). Thus, this vaccine was found to be effective against SARS-CoV-2 as well because both viruses use the same mechanism to invade host cells.

Gold nanoparticles (AuNPs) in conjugation with transmissible gastroenteritis virus (TGEV), a type of coronavirus, were studied to check immunity in mice and rabbits (Staroverov et al. 2011). It was observed that there was immunization against TGEV due to an increase in antibody titer, the concentration of IFN- γ, tenfold expansion of T cells, and macrophage respiratory activity (Jafari et al. 2022). The vaccine made with AuNPs-adjuvanted S protein was able to produce an IgG immune response and was found to be effective against SARS-CoV2 infection (Sekimukai et al. 2020). Figure 14.8 shows the various COVID-19 vaccines available in the market in accordance with their vaccine platform.

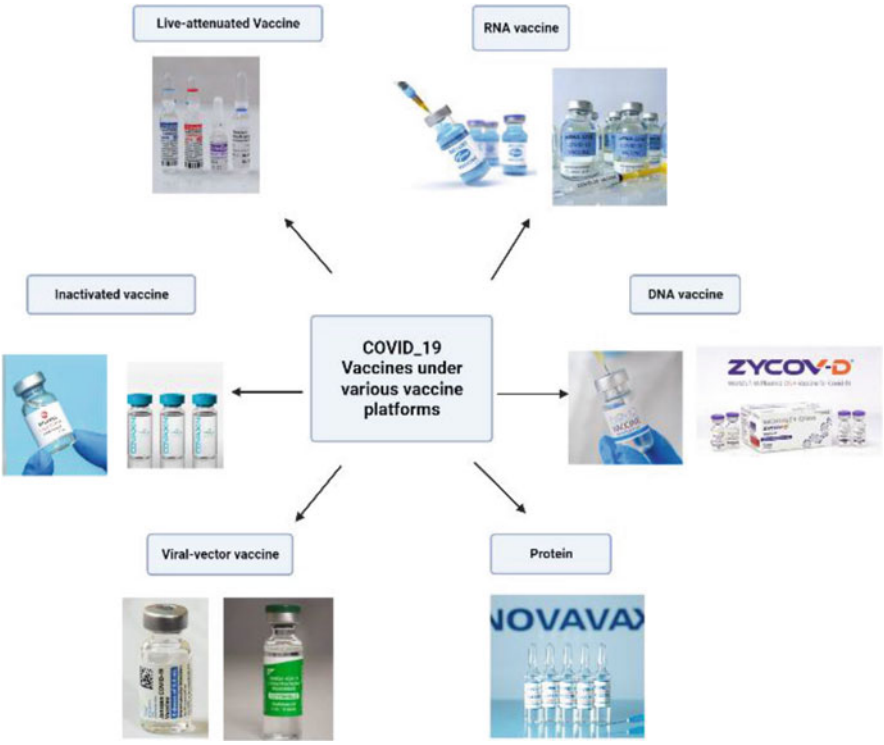

Fig. 14.8 Different vaccines so far available for COVID-19

14.5 Use of New Generation Vaccine Technology and Reverse Vaccinology to Control Pandemics in Animals

Vaccination is the most effective way to prevent or overcome the pandemic. This has been reiterated during the COVID-19 pandemic, wherein, vaccines have been developed most rapidly in the history of medical science using different vaccine technologies as discussed above. Different platform technologies elicit distinct types of immune responses in the host body. The information of such technologies, together with knowledge on pathogen, host pathogen–interaction, and immune response induced by the pathogen can be used to design novel veterinary vaccines for the diseases that potentially may result into pandemics. The information thus available for veterinary pathogens together with reverse vaccinology techniques can be exploited to design novel and effective vaccines for the veterinary pathogens capable of creating pandemics situations, such as swine influenza, avian influenza, etc.

Biotechnological advances have improved vaccine development technology. Further, research has been enhanced in the field of "omics." As a result, now we

have significant data available in databases in terms of genomics, proteomics, and transcriptomics for veterinary pathogens. The advancements in biotechnology, sequencing technologies, and developments in the "omics" world have introduced a "third generation" of vaccines that are developed using reverse vaccinology (Odir et al. 2011; Rappuoli et al. 2014).

Genomic analyses and novel research have resulted in better comprehension of host–pathogen interactions and mechanisms of pathogenesis. The information, thus acquired, has aided in recognizing better antigens as vaccine candidates and improvised recombinant veterinary vaccines. Advancement in sequencing technologies has resulted in whole-genome sequencing of various viruses, and pathogenic eukaryotes and prokaryotes (Kremer et al. 2016; Pizza et al. 2000). These developments have contributed to a better characterization of pathogens, their antigens, and variability among various strains (Aurrecoechea et al. 2007; Cho et al. 2002). With the help of genome sequencing and computational technology, the genome and proteome of pathogens can be studied and screened. The relevant antigenic structures of the pathogens that could elicit the best potent immune response can be identified and exploited for either development of recombinant vaccines or improvising existing subunit vaccines (Seib et al. 2012). Further, screening of novel vaccine candidate antigens can be done in genomic databases of pathogens, wherein the whole genome with complete information on encoded proteins is available (Bagnoli et al. 2011). In general, bacterial surface antigens, secreted proteins, and toxins have been proved to be potent vaccine candidates that elicit sufficient immune response (Ravipaty and Reilly 2010). Additionally, bioinformatics tools can be exploited to study the gene sequences in detail. Gene comparative analysis can be done using the comparative genomic analysis software. Open reading frames carrying genes encoding proteins with known functions can be studied in genomic or proteomic databases. The degree of conservation can be studied among related populations of pathogens using comparative genomic analysis tools. The genes or proteins that are conserved among larger populations of pathogens prove to be good vaccine candidates. Further, improved and careful vaccine design using reverse vaccinology and recombinant technology for the candidate antigens will provide enhanced immune response against pathogens that are antigenically distinct (Seib et al. 2012). These genes thus can be identified in silico, without growing the microorganism in the laboratory. The genes are expressed in mammalian or nonmammalian expression systems such as yeast, insect, *E. coli*, etc.. Expressed proteins are purified and formulated into vaccine form. The vaccine is then administered to the host to elicit an immune response.

The avenues of recombinant protein expression in plants offer edible vaccines which could be a convenient alternative to conventional vaccines (Jorge and Dellagostin 2017). Further, higher expression levels may be obtained for such gene sequences by using recombinant technology. The next-generation vaccines produced using newer technologies are purified forms of recombinant proteins, could be multivalent, have regulated safety profiles, induce immunity, and are better alternatives to conventional vaccines (Oliveira et al. 2015; Rappuoli 2001). These vaccines, however, may exhibit lower immunogenicity as compared to conventional

vaccines. The low immunogenicity could be attributed to the inclusion of only a few antigens as vaccines. The use of potent adjuvants in vaccine formulations may overcome this problem.

The field of veterinary science has been successful in controlling the past epidemics such as bluetongue, foot and mouth disease, bovine pleura-pneumonia, etc. Further, the zoonotic diseases such as avian and swine flu, brucellosis, tuberculosis, bovine spongiform encephalitis, etc., have also been significantly addressed and epidemics have been controlled by the Veterinarians (Ferri and Lloyd-Evans 2021). The diseases with zoonotic potential pose threat to human life as well, and need to be addressed under one health program. Earlier, nearly 70% of infectious diseases such as Ebola, Zika, Nipah encephalitis, and other pandemics such as influenza, SARS, MERS, and HIV/AIDS have originated from wildlife animals (Daszak et al. 2020).

14.6 Adjuvants

A low immunogenicity is usually found associated with recombinant antigens ascribed to the lack of exogenous immune-activating components. Hence, these antigens need to administered with various adjuvants (Soema et al. 2015). Adjuvants trigger the necessary innate immune response needed to optimize the adaptive responses and promote the uptake of vaccine antigens by APCs by two ways: they trigger the innate immune responses that provide a stimulus for antigen presentation or they deliver the antigen in a form already optimized or dendritic cell processing and antigen presentation (Tizard 2020). There are numerous advantages of adjuvants such as increased efficacy, broadening of immune response, and dose sparing to achieve increased availability. Considerable adjuvants have been assessed for use in the field of veterinary vaccines. Major types of adjuvants include aluminum and calcium salts, emulsions (oil-in-water, water-in-oil, and water-in-oil-in-water), liposomes, and archaeosomes, nanoparticles and microparticles, saponins, ISCOMs, polymers, derivatized polysaccharides, carrier proteins, and cytokines (Spickler and Roth 2003). Mineral salt-based adjuvants were the first ones used in vaccines that induced high-titer IgG against extracellular pathogens; however, they have limited applications in vaccine targeting intracellular bacterial infections. ISCOM-based vaccines elicit cellular responses and Th-1-type-1 immunity by IFN-γ production in horses (Paillot et al. 2008); liposomal formulations are known to elicit cytotoxic $CD8^+$ T cell response in pigs (Overgaard et al. 2017); and water-in-oil emulsions elicit Th-1-type immunity by production of IFN-γ in sheep (Begg et al. 2019), which are not seen in the case of mineral salt-based adjuvants. Nanoparticles, cytokines, and polyphosphazenes have not been used in licensed products and are still being evaluated experimentally. Huge interest in alternatives to oil emulsions and mineral salt adjuvants has led to increasing availability of new adjuvants, and hence a number of new adjuvants are in clinical trials.

14.7 Conclusion

The grappling of humankind worldwide into the pandemic of COVID-19 reflects the ultimate need to conceive a promising approach for novel veterinary vaccines, as a means of controlling new or reemerging pathogens. There has been a significant acceleration in the biotechnological advancements over the last two decades, leading to various favorable approaches in the field of vaccinology. The emerging field of genomics has drifted the conventional vaccinology approaches toward gene/sequence-based approaches. Next-generation vaccines have commenced toward increased impact on animal health; however, much is yet to be explored in the field of veterinary vaccines to develop potent, safer, and better characterized animal-specific vaccines. Research may be oriented toward the areas that are yet to be explored. The area of host–pathogen interactions in terms of immune responses in veterinary species remains to be explored in detail. Further, more research needs to be focused in the fields of "omics" and the data together with next-generation vaccine technologies may be explored to devise novel and safer vaccines to control potential pandemics in animals.

References

Agnandji ST, Lell B, Fernandes JF, Abossolo BP, Benjamin M, AL K et al (2012) A phase 3 trial of RTS,S/AS01 malaria vaccine in African infants. N Engl J Med 367(24):2284–2295. https://doi.org/10.1056/nejmoa1208394

Aida V, Pliasas VC, Neasham PJ, North JF, McWhorter KL, Glover SR, Kyriakis CS (2021) Novel vaccine technologies in veterinary medicine: a herald to human medicine vaccines. Front Vet Sci 8(654289):1–20. https://doi.org/10.3389/fvets.2021.654289

Al-Halifa S, Gauthier L, Arpin D, Bourgault S, Archambault D (2019) Nanoparticle-based vaccines against respiratory viruses. Front Immunol 10(22):1–11. https://doi.org/10.3389/fimmu.2019.00022

Al-Kassmy J, Pedersen J, Kobinger G (2020) Vaccine candidates against coronavirus infections. Where does COVID-19 stand? Viruses 12(8):1–18. https://doi.org/10.3390/v12080861

Arimitsu H, Lee JC, Sakaguchi Y, Hayakawa Y, Hayashi M, Nakaura M, Takai H, Lin SN, Mukamoto M, Murphy T, Oguma K (2004) Vaccination with recombinant whole heavy chain fragments of clostridium botulinum type C and D neurotoxins. Clin Diagn Lab Immunol 11(3):496–502. https://doi.org/10.1128/CDLI.11.3.496-502.2004

Aurrecoechea C, Heiges M, Wang H, Wang Z, Fischer S, Rhodes P, Miller J, Kraemer E, Stoeckert CJ Jr, Roos DS, Kissinger JC (2007) ApiDB : integrated resources for the apicomplexan bioinformatics resource center. Nucleic Acids Res 35(Suppl. 1):427–430. https://doi.org/10.1093/nar/gkl880

Babiuk LA, Pontarollo R, Babiuk S, Loehr B, Littel-van VD, den Hurk S (2003) Induction of immune responses by DNA vaccines in large animals. Vaccine 21(7–8):649–658. https://doi.org/10.1016/S0264-410X(02)00574-1

Bagnoli F, Baudner B, Mishra RPN, Bartolini E, Fiaschi L, Mariotti P, Nardi-Dei V, Boucher P, Rappuoli R (2011) Designing the next generation of vaccines for global public health. OMICS: J Integr Biol 15(9):545–566. https://doi.org/10.1089/omi.2010.0127

Begg DJ, Dhungyel O, Naddi A, Dhand NK, Plain KM, de Silva K, Purdie AC, Whittington RJ (2019) The immunogenicity and tissue reactivity of Mycobacterium avium subsp

paratuberculosis inactivated whole cell vaccine is dependent on the adjuvant used. Heliyon 5(6): e01911. https://doi.org/10.1016/j.heliyon.2019.e01911
Boigard H, Alimova A, Martin GR, Katz A, Gottlieb P, Galarza JM (2017) Zika virus-like particle (VLP) based vaccine. PLoS Negl Trop Dis 11(5):1–20. https://doi.org/10.1371/journal.pntd.0005608
Calina D, Docea AO, Petrakis D, Egorov AM, Ishmukhametov AA, Gabibov AG, Shtilman MI, Kostoff R, Carvalho F, Vinceti M, Spandidos DA, Tsatsakis A (2020) Towards effective COVID-19 vaccines: updates, perspectives and challenges (review). Int J Mol Med 46(1): 3–16. https://doi.org/10.3892/ijmm.2020.4596
Calvo-Pinilla E, Marín-López A, Moreno S, Lorenzo G, Utrilla-Trigo S, Jiménez-Cabello L, Benavides J, Nogales A, Blasco R, Brun A, Ortego J (2020) A protective bivalent vaccine against Rift Valley fever and bluetongue. Npj Vaccines 5(1):1–12. https://doi.org/10.1038/s41541-020-00218-y
CDC (2018) Recombivax HB Hepatitis B Vaccine (Recombinant), pp 1–10. https://www.fda.gov/files/vaccines%2Cblood%26biologics/published/package-insert-recombivax-hb.pdf
Chackeriam B (2007) Virus-like particles: flexible platforms for vaccine development. Expert Rev Vaccines 6(3):381–390. https://doi.org/10.1586/14760584.6.3.381
Chapman R, Rybicki EP (2019) Use of a novel enhanced DNA vaccine vector for preclinical virus vaccine investigation. Vaccine 7(2):1–12. https://doi.org/10.3390/vaccines7020050
Chauhan G, Madou MJ, Kalra S, Chopra V, Ghosh D, Martinez-Chapa SO (2020) Nanotechnology for COVID-19: therapeutics and vaccine research. ACS Nano 14(7):7760–7782. https://doi.org/10.1021/acsnano.0c04006
Cho HW, Howard CR, HW, L. (2002) Review of an inactivated vaccine against hantaviruses. Intervirology 45(4–6):328–333. https://doi.org/10.1159/000067925
Crisci E, Bárcena J, Montoya M (2012) Virus-like particles: the new frontier of vaccines for animal viral infections. Vet Immunol Immunopathol 148(3–4):211–225. https://doi.org/10.1016/j.vetimm.2012.04.026
Dai L, Gao GF (2021) Viral targets for vaccines against COVID-19. Nat Rev Immunol 21(2): 73–82. https://doi.org/10.1038/s41577-020-00480-0
De Gregorio E, Rappuoli R (2014) From empiricism to rational design: a personal perspective of the evolution of vaccine development. Nat Rev Immunol 14(7):505–514. https://doi.org/10.1038/nri3694
Del Giudice G, Rappuoli R, Didierlaurent AM (2018) Correlates of adjuvanticity: a review on adjuvants in licensed vaccines. Semin Immunol 39:14–21. https://doi.org/10.1016/j.smim.2018.05.001
Deml L, Speth C, Dierich MP, Wolf H, Wagner R (2005) Recombinant HIV-1 Pr55gag virus-like particles: potent stimulators of innate and acquired immune responses. Mol Immunol 42(2): 259–277. https://doi.org/10.1016/j.molimm.2004.06.028
Dhanasooraj D, Kumar RA, Mundayoor S (2016) Subunit protein vaccine delivery system for tuberculosis based on hepatitis B virus Core VLP (HBc-VLP) particles. Methods Mol Biol 1404:377–392. https://doi.org/10.1007/978-1-4939-3389-1_26
Dhanwani R, Ly H, Liang Y (2017) Recombinant tri-segmented Pichinde virus as a novel live viral vaccine platform. Methods Mol Biol 1581:169–179. https://doi.org/10.1007/978-1-4939-6869-5_10
Diamond MS, Pierson TC (2020) The challenges of vaccine development against a new virus during a pandemic. Cell Host Microbe 27(5),:699–703. https://doi.org/10.1016/j.chom.2020.04.021
Dintzis RZ (1992) Rational design of conjugate vaccines. Pediatr Res 32(4):376–385. https://doi.org/10.1203/00006450-199210000-00002
Dormont F, Brusini R, Cailleau C, Reynaud F, Reynaud F, Peramo A, Gendron A, Mougin J, Gaudin F, Gaudin F, Varna M, Couvreur P (2020) Squalene-based multidrug nanoparticles for improved mitigation of uncontrolled inflammation in rodents. Sci Adv 6(23):1–11. https://doi.org/10.1126/sciadv.aaz5466

El-Attar L, Oliver SL, Mackie A, Charpilienne A, Poncet D, Cohen J, Bridger JC (2009) Comparison of the efficacy of rotavirus VLP vaccines to a live homologous rotavirus vaccine in a pig model of rotavirus disease. Vaccine 27(24):3201–3208. https://doi.org/10.1016/j.vaccine.2009.03.043

Entrican G, Francis MJ (2022) Applications of platform technologies in veterinary vaccinology and the benefits for one health. Vaccine 40(20):2833–2840. https://doi.org/10.1016/j.vaccine.2022.03.059

Eshghi A, Cullen PA, Cowen L, Zuerner RL, Cameron CE (2009) Global proteome analysis of Leptospira interrogans. J Proteome Res 8(10):4564–4578. https://doi.org/10.1021/pr9004597

Esmaeilzadeh A, Elahi R (2021) Immunobiology and immunotherapy of COVID-19: a clinically updated overview. J Cell Physiol 236(4):2519–2543. https://doi.org/10.1002/jcp.30076

Ferraro B, Morrow MP, Hutnick NA, Shin TH, Lucke CE, Weiner DB (2011) Clinical applications of DNA vaccines: current progress. Clin Infect Dis 53(3):296–302. https://doi.org/10.1093/cid/cir334

Ferri M, Lloyd-Evans M (2021) The contribution of veterinary public health to the management of the COVID-19 pandemic from a one health perspective. One Health 12:100230. https://doi.org/10.1016/j.onehlt.2021.100230

Francis MJ (2018) Recent advances in vaccine technologies. Vet Clin North Am Small Anim Pract 48(2):231–241. https://doi.org/10.1016/j.cvsm.2017.10.002

Franco D, Liu W, Gardiner DF, Hahn BH, Ho DD (2011) CD40L-containing virus-like particle as a candidate HIV-1 vaccine targeting dendritic cells. J Acquir Immune Defic Syndr 56(5):393–400. https://doi.org/10.1097/QAI.0b013e31820b844e

Frederiksen LSF, Zhang Y, Foged C, Thakur A (2020) The long road toward COVID-19 herd immunity: vaccine platform technologies and mass immunization strategies. Front Immunol 11(1817):1–26. https://doi.org/10.3389/fimmu.2020.01817

Frietze KM, Peabody DS, Chackerian B (2016) Engineering virus-like particles as vaccine platforms. Curr Opin Virol 18:44–49. https://doi.org/10.1016/j.coviro.2016.03.001

Ghosh S, Nagar G (2014) Problem of ticks and tick-borne diseases in India with special emphasis on progress in tick control research: a review. J Vector Borne Dis 51(4):259–270

GlaxoSmithKline (1989) ENGERIX-B is indicated for immunization against infection caused by all known subtypes of hepatitis B virus. For intramuscular administration. See Section 2.2 for subcutaneous administration in persons at risk of hemorrhage. Preparation for Administr. 1–16. https://www.fda.gov/media/119403/download

GlaxoSmithKline (2009) CERVARIX [Human Papillomavirus Bivalent (Types 16 and 18) Vaccine, Recombinant] Suspension for Intramuscular injection. 1–28. https://www.fda.gov/media/78013/download

Grgacic EVL, Anderson DA (2006) Virus-like particles: Passport to immune recognition. Methods 40(1):60–65. https://doi.org/10.1016/j.ymeth.2006.07.018

Han Y, Král P (2020) Computational design of ACE2-based peptide inhibitors of SARS-CoV-2. ACS Nano 14(4):5143–5147. https://doi.org/10.1021/acsnano.0c02857

Han J, Zhao D, Li D, Wang X, Jin Z, Zhao K (2018) Polymer-based nanomaterials and applications for vaccines and drugs. Polymers 10(1):1–14. https://doi.org/10.3390/polym10010031

Hartwig DD, Oliveira TL, Seixas FK, Forster KM, Rizzi C, Hartleben CP, Mcbride AJA, Dellagostin OA (2010) High yield expression of leptospirosis vaccine candidates LigA and LipL32 in the methylotrophic yeast Pichia. Pastoris 9(98):1–7

Heinson AI, Woelk CH, Newell ML (2015) The promise of reverse vaccinology. Int Health 7(2):85–89. https://doi.org/10.1093/inthealth/ihv002

Hobernik D, Bros M (2018) DNA vaccines — how far from clinical use ? Figure 1:19(3605),1–28. https://doi.org/10.3390/ijms19113605

Hu CMJ, Chang WS, Fang ZS, Chen YT, Wang WL, Tsai HH, Chueh LL, Takano T, Hohdatsu T, Chen HW (2017) Nanoparticulate vacuolar ATPase blocker exhibits potent host-targeted antiviral activity against feline coronavirus. Sci Rep 7(1):1–11. https://doi.org/10.1038/s41598-017-13316-0

IPBES (2020) Workshop report on biodiversity and pandemics of the intergovernmental platform on biodiversity and ecosystem services. Daszak, P., Amuasi, J., das Neves, C. G., Hayman, D., Kuiken, T., Roche, B., Zambrana-Torrelio, C., Buss, P., Dundarova, H., Feferholtz, Y., Földvári, G., Igbinosa, E., Junglen, S., Liu, Q., Suzan, G., Uhart, M., Wannous, C., Woolaston, K., Mosig Reidl, P., O'Brien, K., Pascual, U., Stoett, P., Li, H., Ngo, H. T., IPBES secretariat, Bonn, Germany, DOI:https://doi.org/10.5281/zenodo.4147317

Jafari A, Danesh Pouya F, Niknam Z, Abdollahpour-Alitappeh M, Rezaei-Tavirani M, Rasmi Y (2022) Current advances and challenges in COVID-19 vaccine development: from conventional vaccines to next-generation vaccine platforms. Mol Biol Rep 49:4943. https://doi.org/10.1007/s11033-022-07132-7

Jeyanathan M, Afkhami S, Smaill F, Miller MS, Lichty BD, Xing Z (2020) Immunological considerations for COVID-19 vaccine strategies. Nat Rev Immunol 20(10):615–632. https://doi.org/10.1038/s41577-020-00434-6

Jorge S, Dellagostin OA (2017) The development of veterinary vaccines: a review of traditional methods and modern biotechnology approaches. Biotechnol Res Innov 1(1):6–13. https://doi.org/10.1016/j.biori.2017.10.001

Khan KH (2013) DNA vaccines: roles against diseases. Germs 3(1):26–35. https://doi.org/10.11599/germs.2013.1034

Kremer FS, Eslabão MR, Jorge S, Oliveira NR, Labonde J, Santos MNP, Monte LG, Grassmann AA, Cunha CEP, Forster KM, Moreno LZ, Moreno AM, Campos VF, McBride AJA, Pinto LS, Dellagostin OA (2016) Draft genome of the leptospira interrogans strains, acegua, RCA, prea, and capivara, obtained from wildlife maintenance hosts and infected domestic animals. Mem Inst Oswaldo Cruz 111(4):280–283. https://doi.org/10.1590/0074-02760160010

Kushnir N, Streatfield SJ, Yusibov V (2012) Virus-like particles as a highly efficient vaccine platform: diversity of targets and production systems and advances in clinical development. Vaccine 31(1):58–83. https://doi.org/10.1016/j.vaccine.2012.10.083

Leung AKC (2011) "Variolation" and vaccination in late imperial China. In: Pltokin SA (ed) History of vaccine development. Springer, pp 5–12. https://doi.org/10.1007/978-1-4419-1339-5

Li H, Li Y, Wang X, Hou Y, Hong X, Gong T, Zhang Z, Sun X (2017a) Rational design of polymeric hybrid micelles to overcome lymphatic and intracellular delivery barriers in cancer immunotherapy. Theranostics 7(18):4383. https://doi.org/10.7150/thno.20745

Li H, Shao S, Cai J, Burner D, Lu L, Chen Q, Minev B, Ma W (2017b) Artificial human antigen-presenting cells are superior to dendritic cells at inducing cytotoxic T-cell responses. Immunology 152(3):462–471. https://doi.org/10.1111/imm.12783

Maina TW, Grego EA, Boggiatto PM, Sacco RE, Narasimhan B, McGill JL (2020) Applications of Nanovaccines for disease prevention in cattle. Front Bioeng Biotechnol 8:1–20. https://doi.org/10.3389/fbioe.2020.608050

Meeusen ENT, Walker J, Peters A, Pastoret PP, Jungersen G (2007) Current status of veterinary vaccines. Clin Microbiol Rev 20(3):489–510. https://doi.org/10.1128/CMR.00005-07

Melnick JL (1978) Advantages and disadvantages of killed and live poliomyelitis vaccines. Bull World Health Organ 56(1):21–38

Mendonça SA, Lorincz R, Boucher P, Curiel DT (2021) Adenoviral vector vaccine platforms in the SARS-CoV-2 pandemic. Npj Vaccines 6(1):97. https://doi.org/10.1038/s41541-021-00356-x

Merad M, Martin JC (2020) Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. Nat Rev Immunol 20(6):355–362. https://doi.org/10.1038/s41577-020-0331-4

Merck&Co (2015) Package insert GARDASIL® [human papillomavirus Quadrivalent (types 6, 11, 16, and 18) vaccine, recombinant] suspension for intramuscular injection. Merck & Co, pp 1–28. https://www.merck.com/product/usa/pi_circulars/g/gardasil/gardasil_pi.pdf, https://www.fda.gov/files/vaccines,blood&biologics/published/Package-Insert%2D%2D-Gardasil.pdf

Mohsen MO, Zha L, Cabral-Miranda G, Bachmann MF (2017) Major findings and recent advances in virus–like particle (VLP)-based vaccines. Semin Immunol 34:123–132. https://doi.org/10.1016/j.smim.2017.08.014

Moon JJ, Suh H, Li AV, Ockenhouse CF, Yadava A, Irvine DJ (2012) Enhancing humoral responses to a malaria antigen with nanoparticle vaccines that expand T fh cells and promote germinal center induction. Proc Natl Acad Sci U S A 109(4):1080–1085. https://doi.org/10.1073/pnas.1112648109

Moreira C, da Cunha CEP, Moreira GMSG, Mendonça M, Salvarani FM, Moreira ÂN, Conceição FR (2016) Protective potential of recombinant non-purified botulinum neurotoxin serotypes C and D. Anaerobe 40:58–62. https://doi.org/10.1016/j.anaerobe.2016.05.012

Muhammad W, Zhai Z, Gao C (2020) Antiviral activity of nanomaterials against coronaviruses. Macromol Biosci 20(10):2000196. https://doi.org/10.1002/mabi.202000196

Murata K, Lechmann M, Qiao M, Gunji T, Altert HJ, Liang TJ (2003) Immunization with hepatitis C virus-like particles protects mice from recombinant hepatitis C virus-vaccinia infection. Proc Natl Acad Sci U S A 100(11):6753–6758. https://doi.org/10.1073/pnas.1131929100

Naskalska A, Dabrowska A, Nowak P, Szczepanski A, Jasik K, Milewska A, Ochman M, Zeglen S, Rajfur Z, Pyrc K (2018) Novel coronavirus-like particles targeting cells lining the respiratory tract. PLoS One 13(9):1–21. https://doi.org/10.1371/journal.pone.0203489

Netland J, DeDiego ML, Zhao J, Fett C, Álvarez E, Nieto-Torres JL, Enjuanes L, Perlman S (2010) Immunization with an attenuated severe acute respiratory syndrome coronavirus deleted in E protein protects against lethal respiratory disease. Virology 399(1):120–128. https://doi.org/10.1016/j.virol.2010.01.004

Odir AD, Grassmann AA, Hartwig DD, Félix SR, Da Silva ÉF, McBride AJA (2011) Recombinant vaccines against leptospirosis. Hum Vaccin 7(11):1215–1224. https://doi.org/10.4161/hv.7.11.17944

Oliveira TL, Grassmann AA, Schuch RA, Neto ACPS, Mendonça M, Hartwig DD, McBride AJA, Dellagostin OA (2015) Evaluation of the leptospira interrogans outer membrane protein OmpL37 as a vaccine candidate. PLoS One 10(11):1–13. https://doi.org/10.1371/journal.pone.0142821

Overgaard NH, Frøsig TM, Jakobsen JT, Buus S, Andersen MH, Jungersen G (2017) Low antigen dose formulated in CAF09 adjuvant Favours a cytotoxic T-cell response following intraperitoneal immunization in Göttingen minipigs. Vaccine 35(42):5629–5636. https://doi.org/10.1016/j.vaccine.2017.08.057

Paillot R, Grimmett H, Elton D, Daly JM (2008) Protection, systemic IFNγ, and antibody responses induced by an ISCOM-based vaccine against a recent equine influenza virus in its natural host. Vet Res 39(3):21. https://doi.org/10.1051/vetres:2007062

Paliard X, Liu Y, Wagner R, Wolf H, Baenziger J, Walker CM (2000) Priming of strong, broad, and long-lived HIV type 1 p55gag-specific CD8+ cytotoxic T cells after Administration of a Virus-like Particle Vaccine in rhesus macaques. AIDS Res Hum Retrovir 16(3):273–282. https://doi.org/10.1089/088922200309368

Panicali D, Paoletti E (1982) Construction of poxviruses as cloning vectors: insertion of the thymidine kinase gene from herpes simplex virus into the DNA of infectious vaccinia virus. Proc Natl Acad Sci U S A 79(16):4927–4931. https://doi.org/10.1073/pnas.79.16.4927

Pati R, Shevtsov M, Sonawane A (2018) Nanoparticle vaccines against infectious diseases. Front Immunol 9:2224. https://doi.org/10.3389/fimmu.2018.02224

Pillay S, Shephard EG, Meyers AE, Williamson AL, Rybicki EP (2010) HIV-1 sub-type C chimaeric VLPs boost cellular immune responses in mice. J Immune Based Ther Vaccines 8(7):2–7. https://doi.org/10.1186/1476-8518-8-7

Pimentel TAPF, Yan Z, Jeffers SA, Holmes KV, Hodges RS, Burkhard P (2009) Peptide nanoparticles as novel immunogens: design and analysis of a prototypic severe acute respiratory syndrome vaccine. Chem Biol Drug Des 73(1):53–61. https://doi.org/10.1111/j.1747-0285.2008.00746.x

Pizza M, Scarlato V, Masignani V, Giuliani M, Aricò B, Comanducci M, Jennings G, Baldi L, Bartolini E, Capecchi B, Galeotti C, Luzzi E, Manetti R, Marchetti E, Mora M, Nuti S, Ratti G, Santini L, Savino S, Rappuoli R (2000) Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. Science 287(5459):1816–1820. https://doi.org/10.1126/science.287.5459.1816

Plotkin S (2014) History of vaccination. Proc Natl Acad Sci U S A 111(34):12283–12287. https://doi.org/10.1073/pnas.1400472111

Rappuoli R (2001) Reverse vaccinology, a genome-based approach to vaccine development. Vaccine 19(17–19):2688–2691. https://doi.org/10.1016/S0264-410X(00)00554-5

Rappuoli R, Pizza M, Del Giudice G, De Gregorio E (2014) Vaccines, new opportunities for a new society. Proc Natl Acad Sci U S A 111(34):12288–12293. https://doi.org/10.1073/pnas.1402981111

Ravipaty S, Reilly JP (2010) Comprehensive characterization of methicillin-resistant Staphylococcus aureus subsp. aureus COL secretome by two-dimensional liquid chromatography and mass spectrometry. Mol Cell Proteomics 9(9):1898–1919. https://doi.org/10.1074/mcp.M900494-MCP200

Riedel S (2005) Edward Jenner and the history of smallpox and vaccination. Bayl Univ Med Cent Proc 18(1):21–25. https://doi.org/10.1080/08998280.2005.11928028

Rizzi C, Bianco MV, Blanco FC, Soria M, Gravisaco MJ, Montenegro V, Vagnoni L, Buddle B, Garbaccio S, Delgado F, Leal KS, Cataldi AA, Dellagostin OA, Bigi F (2012) Vaccination with a BCG strain overexpressing Ag85B protects cattle against Mycobacterium bovis challenge. PLoS One 7(12):1–10. https://doi.org/10.1371/journal.pone.0051396

Sahin U, Muik A, Derhovanessian E, Vogler I, Kranz LM, Vormehr M, Baum A, Pascal K, Quandt J, Maurus D, Brachtendorf S, Lörks V, Sikorski J, Hilker R, Becker D, Eller AK, Grützner J, Boesler C, Rosenbaum C, Türeci Ö et al (2020) COVID-19 vaccine BNT162b1 elicits human antibody and TH1 T cell responses. Nature 586(7830):594–599. https://doi.org/10.1038/s41586-020-2814-7

Salian VS, Wright JA, Vedell PT, Nair S, Li C, Kandimalla M, Tang X, Carmona Porquera EM, Kalari KR, Kandimalla KK (2021) COVID-19 transmission, current treatment, and future therapeutic strategies. Mol Pharm 18(3):754–771. https://doi.org/10.1021/acs.molpharmaceut.0c00608

Schirmbeck R, Böhm W, Reimann J (1996) Virus-like particles induce MHC class I-restricted T-cell responses. Lessons learned from the hepatitis B small surface antigen. Intervirology 39(1–2):111–119. https://doi.org/10.1159/000150482

Seib KL, Zhao X, Rappuoli R (2012) Developing vaccines in the era of genomics: a decade of reverse vaccinology. Clin Microbiol Infect 18(SUPPL. 5):109–116. https://doi.org/10.1111/j.1469-0691.2012.03939.x

Sekimukai H, Iwata-Yoshikawa N, Fukushi S, Tani H, Kataoka M, Suzuki T, Hasegawa H, Niikura K, Arai K, Nagata N (2020) Gold nanoparticle-adjuvanted S protein induces a strong antigen-specific IgG response against severe acute respiratory syndrome-related coronavirus infection, but fails to induce protective antibodies and limit eosinophilic infiltration in lungs. Microbiol Immunol 64(1):33–51. https://doi.org/10.1111/1348-0421.12754

Shams H (2005) Recent developments in veterinary vaccinology. Vet J 170(3):289–299. https://doi.org/10.1016/j.tvjl.2004.07.004

Simionatto S, Marchioro SB, Galli V, Hartwig DD, Carlessi RM, Munari FM, Laurino JP, Conceição FR, Dellagostin OA (2010) Cloning and purification of recombinant proteins of mycoplasma hyopneumoniae expressed in Escherichia coli. Protein Expr Purif 69(2):132–136. https://doi.org/10.1016/j.pep.2009.09.001

Soema PC, Kompier R, Amorij JP, Kersten GFA (2015) Current and next generation influenza vaccines: formulation and production strategies. Eur J Pharm Biopharm 94:251–263. https://doi.org/10.1016/j.ejpb.2015.05.023

Song JM, Wang BZ, Park KM, van Rooijen N, Quan FS, Kim MC, Jin HT, Pekosz A, Compans RW, Kang SM (2011) Influenza virus-like particles containing M2 induce broadly cross protective immunity. PLoS One 6(1):e14538. https://doi.org/10.1371/journal.pone.0014538

Spickler AR, Roth JA (2003) Adjuvants in veterinary vaccines: modes of action and adverse effects. J Vet Intern Med 17(3):273–281. https://doi.org/10.1111/j.1939-1676.2003.tb02448.x

Staroverov SA, Vidyasheva IV, Gabalov KP, Vasilenko OA, Laskavyi VN, Dykman LA (2011) Immunostimulatory effect of gold nanoparticles conjugated with transmissible gastroenteritis virus. Bull Exp Biol Med 151(4):436–439. https://doi.org/10.1007/s10517-011-1350-8

Suarez DL (2012) DIVA vaccination strategies for avian influenza virus. Avian Dis 56(4s1): 836–844. https://doi.org/10.1637/10207-041512-Review.1

Sun SH, Chen Q, Gu HJ, Yang G, Wang YX, Huang XY, Liu SS, Zhang NN, Li XF, Xiong R, Guo Y, Deng YQ, Huang WJ, Liu Q, Liu QM, Shen YL, Zhou Y, Yang X, Zhao TY, Wang YC et al (2020) A mouse model of SARS-CoV-2 infection and pathogenesis. Cell Host Microbe 28(1):124–133.e4. https://doi.org/10.1016/j.chom.2020.05.020

Tiwari A, Gautam A, Bhat S, Malik YS (2020) Advances and applications of vectored vaccines in animal diseases. In: Khurana SMP, Malik YS, Barh D, Azevedo V (eds) *Genomics and biotechnological advances in veterinary, poultry, and fisheries*. Academic press (Elsevier), pp 361–380. https://doi.org/10.1016/B978-0-12-816352-8.00015-1

Tizard IR (2020) Adjuvants and Adjuvanticity. In: Vaccines for Veterinarians, vol ed.1, pp 75–86. https://doi.org/10.1016/B978-0-323-68299-2.00016-2

van Gelder P, Makoschey B (2012) Production of viral vaccines for veterinary use. Berl Munch Tierarztl Wochenschr 125(3–4):103–109. https://doi.org/10.2376/0005-9366-125-103

van Riel D, de Wit E (2020) Next-generation vaccine platforms for COVID-19. Nat Mater 19(8): 810–812. https://doi.org/10.1038/s41563-020-0746-0

Volz A, Sutter G (2017) Modified vaccinia virus Ankara: history, value in basic research, and current perspectives for vaccine development. In: Advances in virus research, vol 97, 1st edn. Elsevier Inc, p 97. https://doi.org/10.1016/bs.aivir.2016.07.001

Wang J, Peng Y, Xu H, Cui Z, Williams RO (2020) The COVID-19 vaccine race: challenges and opportunities in vaccine formulation. AAPS PharmSciTech 21(6):225. https://doi.org/10.1208/s12249-020-01744-7

Win SJ, Ward VK, Dunbar PR, Young SL, Baird MA (2011) Cross-presentation of epitopes on virus-like particles via the MHC I receptor recycling pathway. Immunol Cell Biol 89(6): 681–688. https://doi.org/10.1038/icb.2010.161

Ye S, Shao K, Li Z, Guo N, Zuo Y, Li Q, Lu Z, Chen L, He Q, Han H (2015) Antiviral activity of graphene oxide: how sharp edged structure and charge matter. ACS Appl Mater Interfaces 7(38):21578–21579. https://doi.org/10.1021/acsami.5b06876

Zhao L, Seth A, Wibowo N, Zhao CX, Mitter N, Yu C, Middelberg APJ (2014) Nanoparticle vaccines. Vaccine 32(3):327–337. https://doi.org/10.1016/j.vaccine.2013.11.069

Zhou H, Guo L, Wang M, Qu J, Zhao Z, Wang J, Hung T (2011) Prime immunization with rotavirus VLP 2/6 followed by boosting with an adenovirus expressing VP6 induces protective immunization against rotavirus in mice. Virol J 8:2–9. https://doi.org/10.1186/1743-422X-8-3

Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F (2020) Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell*. Discovery 6(1):14. https://doi.org/10.1038/s41421-020-0153-3

Chapter 15
Point-of-Care Diagnostics Using Molecular Approaches

Vinay G. Joshi, Manish Tiwari, Mahavir Singh, Richa Khirbat, and Sushila Maan

Abstract Molecular point care diagnostics aimed to provide simplified diagnostic platforms which are rapid and designed in miniaturized forms for on-field disease diagnosis. There are various approaches to develop molecular point-of-care tests (POCTs). In this chapter, we reviewed commonly used molecular POCT platforms. The selection of diagnostic module for effective POCT development usually depends on the factors such as simplicity, robustness, sensitivity, and cost of the test. Currently, isothermal amplification-based POCTs are gaining importance for their sensitive and low-cost input nature. While alternative to amplification, non-amplifying molecular POCT modules are also available. Despite the high relevance POCTs in clinical on-field diagnosis, very few molecular POCTs could be commercialized. These narrow success on-field applications could be overcome with an adaptation of appropriate quality assurance protocols. In nutshell, the molecular POCTs could be handy diagnostic support for augmenting livestock health and production provided appropriate diagnostic module with strict quality assurance criteria should be selected.

Keywords POCT · Nucleic acid · Microfluidics · Aptamers · Isothermal amplification · Paper-based amplification · SPR · Biosensor

V. G. Joshi (✉) · M. Tiwari · R. Khirbat · S. Maan
Department of Animal Biotechnology, College of Veterinary Sciences, Lala Lajpat Rai University of Veterinary and Animal Sciences (LUVAS), Hisar, Haryana, India
e-mail: vinaygjoshi18@luvas.edu.in

M. Singh
College Central Laboratory, College of Veterinary Sciences, Lala Lajpat Rai University of Veterinary and Animal Sciences (LUVAS), Hisar, Haryana, India

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_15

15.1 Introduction

Molecular point-of-care testing (POCT) detection platforms are rapid and miniaturized in nature which facilitates on-field diagnosis, monitoring, and management of the animal disease. The point-of-care testing device terminology predominantly refers to the diagnostic tests robust enough to be performed outside the conventional diagnostic laboratory setup and can be applied in close proximity to patients. The term close proximity can cite as bedside, extra-laboratory, or in-clinic; for animal or livestock concern, the proximity may be represented by word "pen-side" which implies animal herd. The purpose of the POCT is to streamline the diagnostic process together with making healthcare faster and more efficient with reduced cost (Fig. 15.1). Any POCT should follow affordable, sensitive, specific, user-friendly, rapid, equipment-free, delivered (ASSURED) criteria established by the world health organization (WHO) (Peeling et al. 2006). According to market analyst report published in 2020, the livestock and animal POCT market size was valued at USD 1.15 billion in 2020 with average estimate of annual growth rate of 11.3% from 2021 to 2028 (Veterinary Point of Care Diagnostics Market Report, 2028 (grandviewresearch.com)). This estimate helps to understand the significance of POCT platforms in animal diagnosis. Broadly, POCT platforms can be classified as

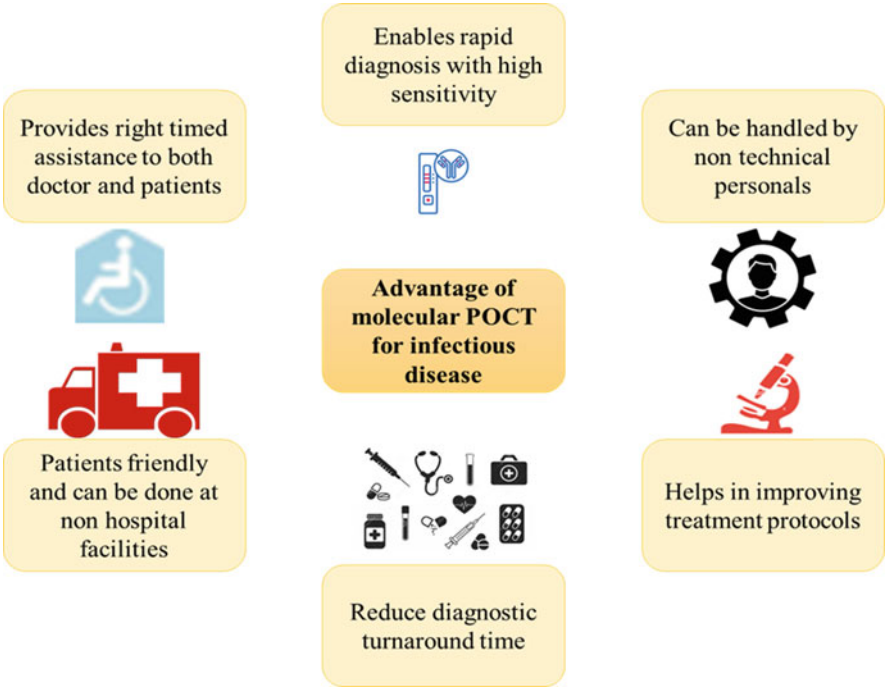

Fig. 15.1 Advantage and scope of the POCTs for clinical diagnosis

traditional or non-molecular POCTs and molecular POCTs. Suboptimal limits of detection and false positivity are some of the limitations of traditional POCTs.

The basic principle of molecular POCTs allows rapid nucleic acid detections. Additionally, molecular POCTs coupled with the amplification of target nucleic acid help in improving diagnostic sensitivity to many folds. The molecular POCTs can be developed using various diagnostic platforms. Recently, the utility of several technologies such as microfluidics, biosensor, SPR, lateral flow assay, PCR, ELISA, aptamer, etc., to develop the rapid diagnostic kit is demonstrated (Vashist 2017; Nayak et al. 2017). These platforms can widely be used for molecular diagnosis of pathogenic diseases, metabolic disorders, cancers, and in the development of clinical biomarkers. The major focus of POCTs in animal and livestock context is rapid, robust diagnostic POCTs with low-cost and pen-side application. Various modules of POCTs are available currently; selection of module depends on various criteria such as target molecules, reporter system, and diagnostic performance. In present book chapter, we will take overview of molecular diagnostic modules of POCT and discuss the various performance parameters to be taken care in terms of quality assurance for improved clinical diagnosis.

15.2 Nucleic Acid Amplification Methods for POCTs

The amplification of target nucleic acid is done using a polymerase enzyme in traditional amplification-based diagnostic process (Fig. 15.2). In the molecular POCT development, isothermal nucleic acid amplification methods are preferred. Isothermal amplification is natural choice in development of POCT platforms for many oblivious reasons such as elimination of thermal cycling steps, reduced instrumentation, increase in sensitivity, and reduction in turnaround time. Commonly used isothermal amplification technologies are as follows: helicase-dependent amplification (HDA), loop-mediated isothermal amplification (LAMP), Nick enzyme amplification reaction (NEAR), recombinase polymerase amplification (RPA), and rolling circle amplification (RCA) (Table 15.1). These amplification tests when coupled with reporter systems result in a molecular POCT.

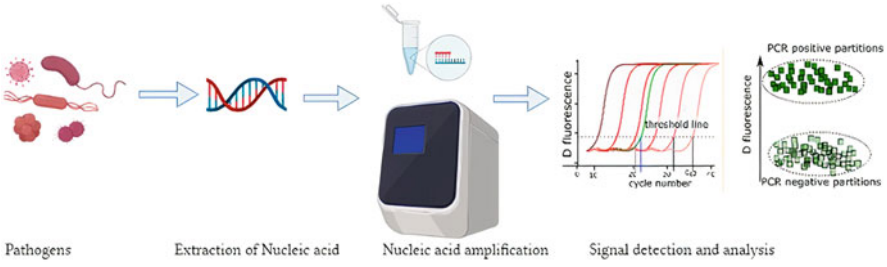

Fig. 15.2 Nucleic acid amplification-based rapid detection of the pathogens

Table 15.1 Isothermal amplification of the nucleic acid

SN	Technique	Component	Multiplex	Examples	References
1	Loop-mediated amplification (LAMP)	Four to six sets of primers, Bst DNA polymerase, dNTPs	Yes	*Plasmodium* spp., *Leishmania, mycobacterium* spp.	Notomi et al. (2015)
2	Recombinase polymerase amplification (RPA)	Two sets of primers, recombinase, SSB, DNA polymerase, REAse Nfo, dNTPs	Yes	*HIV, MRSA, Bacillus subtilis*	Yan et al. (2014)
3	Helicase-dependent amplification (HDA)	Two sets of primers, helicase, SSB, DNA polymerase, dNTPs	Yes	*Ebola virus, Clostridium difficile, Helicobacter pylori, MRSA*	Li and Macdonald (2015)
4	Nucleic acid sequence-based amplification (NASBA)	Two primers, reverse transcriptase, RNase H, RNA polymerase, dNTPs, rNTPs	Yes	*Viruses, Leishmania, plasmodium* spp., *Mycobacterium tuberculosis*	Gill and Ghaemi (2008)
5	Rolling circle amplification (RCA)	Probe, ligase, Phi29 polymerase, dNTPs	NA	*Viruses, Listeria monocytogenes*	Yan et al. (2014)
6	Strand displacement amplification (SDA)	Four primers, Klenow-exo fragment, REase HincII, dGTP, dCTP, dTTP, dATPas	Yes	*Neisseria gonorrhoeae, Chlamydia trachomatis, E. coli*	Li and Macdonald (2015)

15.2.1 Nucleic Acid Detection

POCT of the nucleic acid gives the important information in diagnosis of the infectious disease, genetic biomarker of tumors, and contamination of food and environment. DNA-based diagnostics allow the identification of tumor mutation profiling and highly specific disease markers; however, no genome sequencing is required (Choi et al. 2011). In addition, POCT of nucleic acid directly recognized the pathogens without culturing (Lu et al. 2016). Cordray and Richards-Kortum (2012) reported nucleic acid testing approach for identification of the different species of the *Plasmodium* for malaria diagnosis. In another study, POCT of the nucleic acid for *Mycobacterium tuberculosis* affords a specific and sensitive method of diagnosis (Schito et al. 2015). Hsieh et al. (2022) identified POCT of the nucleic acid test for the detection of the sexually transmitted disease such as chlamydia, gonorrhea, trichomoniasis, and syphilis. Nucleic acid-based molecular diagnostic test requires sequential steps: (1) lysis of the cells and isolating the total DNA or RNA, (2) amplification of the target nucleic acid region, and (3) detection and monitoring of the amplified product. Integration of these three steps resulted in several commercial POC nucleic acid test platforms being available for diagnostic purpose (Niemz et al. 2011; Tables 15.1 and 15.2).

Table 15.2 Commercially available platform for point-of-care nucleic acid testing

SN	Platform	Sample preparation	Amplification	Detection	Time to result (Min)	Examples	References	Website
1	Simple amplification-based assay (SAMBA)	Not required	Isothermal (similar to NASBA)	Nucleic acid lateral flow assay	>60 min	*HIV-1, SARS-CoV-2*	Lee et al. (2010), Assennato et al. (2020)	NA
2	Bioluminescent assay in real- time (BART)	Not required	Isothermal (LAMP)	Real-time fluorescence	< 20	*Chlamydia trachomatis*	Gandelman et al. (2010)	www.optigene.co.uk
3	GeneXpert	Yes	PCR	Real-time fluorescence	<120	*Respiratory syncytial virus, human papillomavirus, HIV, hepatitis B virus*	Gotham et al. (2021)	www.cepheid.com
4	Liat analyzer	Yes	PCR	Real-time fluorescence	<60	*Influenza virus, SARS-CoV-2*	Hansen et al. (2021)	www.iquum.com
5	LA-200	Not required	Isothermal (LAMP)	Real-time turbidimetry	<60	*Mycoplasma pneumoniae*, *Clostridium difficile, human papillomavirus*	Wong et al. (2018)	www.eiken.co.jp
6	Genie II	Not required	Isothermal (LAMP)	Real-time fluorescence	<20	*SARS-CoV-2*	Mautner et al. (2020)	www.optigene.co.uk
7	Twista	Not required	Isothermal (RPA)	Real-time fluorescence	<20	*HIV*	Haleyur Giri Setty and Hewlett (2014)	www.twistdx.co.uk

15.3 Paper-Based Molecular POCTs

Paper-based POCT platforms are cost-effective, rapidly implemented in remote area, portable, and provide easy integration of molecular diagnostic tools. Before the selection of paper module, there is a need to review the inherent properties of paper reflecting capillary flow, porosity, paper thickness, and surface characteristics. These functions play vital role in deciding the analytical fluid flow through paper-based platforms. According to Grant View Research, USA, market share of paper-based diagnostics is about 2.2 billion USD which is expected to reach up to 8.35 billion USD (Websource 2016). Paper-based POCTs can be broadly categorized into three classes namely dipstick, lateral flow assay (LFA), and microfluidic paper-based analytic devices (μPAD). Dipstick molecular POCTs are simplistic diagnostic platforms containing pre-deposited reagents. Several dipstick diagnostic devices are available currently but the major limitation of this approach is it can only provide qualitative detection of analyte (Najian et al. 2016). LFA-based POCTs are modified version of dipstick and provide controlled fluid flow. In general, LFA-based POCT prototypes are made up of four components namely sample pad, conjugation pad, detection pad, and absorbent pad. There are numerous examples of LFA-based molecular diagnostics (Agarwal et al. 2022; Kaur and Toley 2018; Choi et al. 2016; Hu et al. 2014).

The μPAD devices are more robust and address many limitations of other paper-based devices. μPAD devices enable multiplexing and quantitative estimates of analytes (Rezk et al. 2012). Recently, Bender et al. (2021) identified HIV from the human serum samples by using the microfluidic paper-based analytical device (μPAD).

Like other diagnostic tests for clinical samples, paper-based detection of nucleic acid POCT follows the multistep process. The process includes the isolation and purification of the nucleic acid from the clinical samples followed by nucleic acid amplification (using different isothermal amplification method) and detection of the amplified product by tagging them with reporter molecules. In general, paper-based PCOT is an integration of the nucleic acid isolation, amplification, and lateral flow detection. Several diseases such as HIV (Rohrman and Richards-Kortum 2012), chlamydia trachomatis (Linnes et al. 2014), influenza A (Rodriguez et al. 2015), and hepatitis B (Tyas et al. 2021) can be rapidly diagnosed through paper-based biosensor.

15.4 Microfluidic-Based Molecular Diagnostics

POCT by microfluidics facilitates the combination of the process such as extraction of the nucleic acid from the samples, amplification, and detection. Microfluidics implemented different methods for sample preparation to obtain the high quality of nucleic acid such as silicon bead-based extraction (Kim et al. 2016), silicon

micropillar-based biofilter (Petralia et al. 2017), membrane-based extraction (Jangam et al. 2013), paper-based sample preparation (Zhang et al. 2014), and magnetic bead-based nucleic acid extraction (Czilwik et al. 2015). Further, non-isothermal (PCR), isothermal (loop-mediated amplification, recombinase polymerase amplification, helicase-dependent amplification, nucleic acid sequence-based amplification, rolling circle amplification, and strand displacement amplification), and digital PCR are used for amplification of the nucleic acid. Fluorescent and colorimetric signal are generated for the detection of the amplified product. For fluorescence detection, SYBR™ Green, Eva Green™, or SYTO™ Green are the DNA intercalating dyes that can be used to monitor the progress of reaction by measuring the fluorescence emission intensity (Xu et al. 2010). However, leucocrystal violet (LCV) dye can be used for the colorimetric detection (on the basis of color change of reaction) of amplified dsDNA (Miyamoto et al. 2015; Song et al. 2016). Microfluidic-based POCT of nucleic acid is user-friendly, sensitive, quantitative, inexpensive, rapid, and reliable method for diagnosis of the pathogens such as *Salmonella*, *Listeria*, *Cholera,* and *E. coli* (Wang et al. 2021; Nsamela 2020). Chip-based microfluidic device integrated with LAMP can be used for the detection of the waterborne pathogens in 25 min (Jin et al. 2021; Ahmad et al. 2017). In addition, hepatotoxicity could be assessed within a lab-on-a-chip system to analyze the expression of two protein biomarkers by using the multiplex real-time RT-PCR (Lim et al. 2015).

15.5 Nucleic Acid Non-Amplification Methods for POCTs

Simple nucleic acid detection assays without amplifying the target can also provide viable POCT platforms. These methods utilize the molecular interactions between target nucleic acid and reporters. Non-amplification-based systems are easier to handle in terms of field application as they do not require target amplification steps prior diagnostic process. This approach also overcome the limitations of false positivity often encounters in nucleic acid amplification-based POCTs. There are ample examples of such molecular POCT modules which makes use of aptamer probe, peptide nucleic acid (PNA) probe, locked nucleic acid (LNA), or nucleic acid probes (Oh et al. 2017; Saini et al. 2019; Baptista 2018; Joshi et al. 2013).

15.6 Aptamer-Based POCT

Aptamers are short oligonucleotide sequences, generated by SELEX technology and high affinity toward the target molecules. Aptamers based on fluorescence can be used for the detection of the nucleic acid, proteins, amino acid, peptide, antibiotics, and small molecules (Shafiei et al. 2020). Aptamer-based technology was used to diagnosis of certain virus such as influenza (Misono and Kumar 2005) and HIV

(Gopinath 2007). Similarly, Kim et al. (2014) reported the binding of the DNA aptamers to complementary region for the detection of the *E. coli* and Labib et al. (2012) detected the *Salmonella typhimurium* against the aptamers. Recently, aptamer-based technology is used for rapid detection of the SARS–coronavirus (SARS-nCoV19) (Murtaza et al. 2022; Koteswara Rao 2021). Additionally, advancement in the Smartphone-based aptasensors is helpful for the detection of the COVID-19 viruses (Han et al. 2021).

15.7 Gold Nanoparticle-Based Non-Cross-Linking Molecular POCTs

This approach is based on color change in solution resulted from the aggregation of gold nanoprobes. Such aggregation can be achieved through destabilizing nanoprobes in solution with alterations in ionic strength. The outcome of aggregation in the colloidal solution's results to color changes from red to blue. While at the same ionic strength the solution is stabilized when Au-nanoprobes poorly hybridized to a mismatched target resulting in no aggregation and the solution retains its original red color. To conclude, in such scheme change in color is indication of positive sample and retention of color is indication of negative sample. Although this scheme provides a simplistic diagnostic outcome, this method needs a cumbersome optimization of probe activation and combinations of salt concentrations.

Similar principle of color change can be used for molecular POCT development using PNA probes. Naturally, free uncharged PNA probes induce aggregation in gold and silver metallic nanoparticles causing color change while this aggregation induced by PNA is prevented by selective hybridization of PNA probe with the target nucleic acid and color of nanoparticle is retained (Su and Kanjanawarut 2009). This method of non-amplifying nucleic acid with unlabeled and un-conjugated probe-based POCTs is usually referred as label-free POCTs. Using this phenomenon, POCT prototypes were developed for diagnosis of Newcastle disease virus, influenza A virus, and bovine viral diarrhea virus (Joshi et al. 2013; Kumar et al. 2020; Askaravi et al. 2017).

15.8 SPR-Based Nucleic Acid Diagnostics

SPR-based detection of the nucleic acid is simple, highly sensitive, label-free, and real-time detection and requires low amount of sample (Tang and Zhang 2006) Fig. 15.3. POCT by SPR technique was used for the quantitative detection of the avian influenza DNA (Kim et al. 2009). Ding et al. (2015) reported detection of the miRNA by using the SPR biosensor. The SPR-based biosensor could detect very low amount (9pM) of target miRNA in 30 min (Ding et al. 2015). In another study,

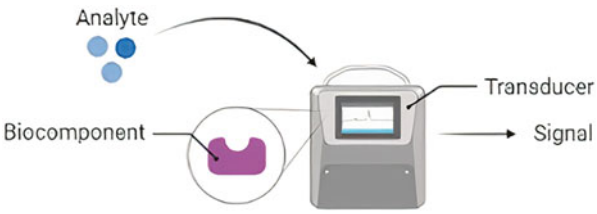

Fig. 15.3 Biosensors workflow

nanograting-based SPR and LAMP method was used for the detection of the hepatitis B virus within 30 min (Lin et al. 2009). Additionally, SPR imaging mode was used for the detection of the viral RNA (Palau et al. 2013). Huang et al. (2020) developed portable and very sensitive SPR sensor for the detection of the nucleic acid (0.01 μmol/mL).

15.9 CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Based Point of Diagnostics

The CRISPR has been widely used for gene editing and transgenic technology. Recently, CRISPR-associated Cas9-based nucleic acid detection provided a groundbreaking finding on diagnostic platform. In the CRISPR-based diagnostic methods, different variant of Cas9 was used to recognize double-stranded DNA (dsDNA). Several robust and effective diagnostic platforms were established by using the CRISPR-Cas9 technology (Pardee et al. 2016). However, the finding of different variant of Cas13a (C2c2) (Shmakov et al. 2015), Cas12a (Cpf1) (Zetsche et al. 2015), CasRx (Konermann et al. 2018), and Cas14 (Harrington et al. 2018), which have collateral cleavage activity has played important role in the nucleic acid detection. Further, the different diagnostic platform has been developed for the detection of the nucleic acid based on these Cas9 proteins. Initially, Gootenberg et al. (2018) developed the diagnostic platform SHERLOCK (specific high-sensitivity enzymatic reporter unlocking) based on the Cas13a, for the detection of the RNA molecule. At the same time, another diagnostic platform, highly efficient system (HOLMES) was developed to exploit the Cas12a (Cheng 2017). The SHERLOCK method integrated with recombinase polymerase amplification (RPA) can be used for the amplification and detection of the viral genetic material (Kocak and Gersbach 2018). Apart from this, another diagnostic platform, HUDSON (Heating Unextracted Diagnostic Samples to Obliterate Nucleases), was developed to detect the viral genetic materials from body fluids such as saliva, blood, urine (Myhrvold et al. 2018). The SHERLOCK and HUDSON methods were applied for diagnosis of the viruses belonging to flaviviruses family such as dengue, West Nile, zika, and yellow fever viruses (Myhrvold et al. 2018). Based on the CRISPR-Cas system, several other diagnostic platforms including STOP Covid, CREST, SHINE, VanGuard, AIOD, and DISCoVER have been developed for the detection of the

COVID-19 (Bhardwaj et al. 2022). Ai et al. (2019) utilized the CRISPR/Cas12 system for the detection of the *Mycobacterium tuberculosis*. Recently, RCA-assisted CRISPR/Cas9 cleavage (RACE) diagnostic platform was used detection of the noncommunicable diseases such as cancer (Wang et al. 2019). Currently, CRISPR/Cas-based platforms have been developed for the detection of the nucleic acid, enzyme, proteins, and contaminants. In future, CRISPR technology integrated with different amplification strategies could be used to develop the diagnostic kit for the detection of the nucleic acid in clinical samples and pathogens.

15.10 Quality Assurance for POCTs

Although various molecular POCTs are developed for livestock and animal disease diagnosis, few of them could comply with farm to land concept. Many of the developed molecular POCTs were successful at laboratory scale and need comprehensive quality assurance checkups. Limited guidelines are available for diagnostic validation of molecular POCTs to see light of the day. The quality assurance criteria for animal POCT diagnostics are comprehensively reviewed by Flatland et al. (2013). The American society for veterinary clinical pathology (ASVCP) through quality assurance and laboratory standards committee (QALS) has established guidelines for quality assurance for POCTs in Veterinary Medicine. The term quality assurance refers to laboratory procedure, which helps in monitoring and improving the diagnostic performance with aim to minimize test error at all stages such as pre-analytical, analytical, and post-analytical. After a diligent literature review, the QALS has recommended several important points for quality assurance of POCTs in Veterinary Medicine, published by Flatland et al. (2013). According to Flatland et al. (2013), major recommendations are (Agarwal et al. 2022) "adapting a formalized approach to POCT within the facility (Ahmad et al. 2017), use and development of written standard operating procedures, policies, forms, and logs (Ai et al. 2019), operator/technicians training, with periodic skills assessment (Askaravi et al. 2017) assessment of instrument analytical performance and regular assessment of statistical quality control (Assennato et al. 2020) use of prevalidated reference intervals (Baptista 2018), ensuring accurate patient results reporting system." In nutshell, the molecular POCTs need to be evaluated thoroughly for their reproducibility and other quality parameters. The guidelines provided by ASVCP could be helpful in improving the clinical diagnostic outcome of various molecular POCTs.

15.11 Conclusion and Future Prospective

The present book chapter focused on the variety of POC diagnosis modules based on various technologies. Nucleic acid-based rapid detection test provides strength to health security, which helps in early clinical diagnosis and control of endemic disease. Recent advancement in POCT offers the simple, user-friendly, sensitive, rapid, reliable, economical, and disposable diagnostic kits. POCTs are improving day by day using different technologies. In the present situation, the most momentous strength for POC diagnostics is the rise of consumer electronics and connected devices. However, the future prospective of POCT depends on improving the affordability of the diagnostic devices.

References

Agarwal S, Warmt C, Henkel J, Schrick L, Nitsche A, Bier FF (2022) Lateral flow–based nucleic acid detection of SARS-CoV-2 using enzymatic incorporation of biotin-labeled dUTP for POCT use. Anal Bioanal Chem Res 414:1–10

Ahmad F, Stedtfeld RD, Waseem H, Williams MR, Cupples AM, Tiedje JM, Hashsham SA (2017) Most probable number-loop mediated isothermal amplification (MPN-LAMP) for quantifying waterborne pathogens in< 25 min. J Microbiol Methods 132:27–33

Ai JW, Zhou X, Xu T, Yang M, Chen Y, He GQ, Pan N, Cai Y, Li Y, Wang X, Su H (2019) CRISPR-based rapid and ultra-sensitive diagnostic test for mycobacterium tuberculosis. Emerg Microbes Infect 8(1):1361–1369

Askaravi M, Rezatofighi SE, Rastegarzadeh S, Shapouri SA, M.R. (2017) Development of a new method based on unmodified gold nanoparticles and peptide nucleic acids for detecting bovine viral diarrhea virus-RNA. AMB Express 7(1):1–9

Assennato SM, Ritchie AV, Nadala C, Goel N, Tie C, Nadala LM, Zhang H, Datir R, Gupta RK, Curran MD, Lee HH (2020) Performance evaluation of the SAMBA II SARS-CoV-2 test for point-of-care detection of SARS-CoV-2. J Clin Microbiol 59(1):e01262–e01220

Baptista PV (2018) Gold nanoprobe-based non-crosslinking hybridization for molecular diagnostics: an update. Expert Rev Mol Diagn 18(9):767–773

Bender AT, Sullivan BP, Zhang JY, Juergens DC, Lillis L, Boyle DS, Posner JD (2021) HIV detection from human serum with paper-based isotachophoretic RNA extraction and reverse transcription recombinase polymerase amplification. Analyst 146(9):2851–2861

Bhardwaj P, Kant R, Behera SP, Dwivedi GR, Singh R (2022) Next-generation diagnostic with CRISPR/Cas: beyond nucleic acid detection. Int J Mol Sci 23(11):6052

Cheng QX (2017) An application of a Cas protein, and a method and kit for detecting a target nucleic acid molecule

Choi S, Goryll M, Sin LYM, Wong PK, Chae J (2011) Microfluidic-based biosensors toward point-of-care detection of nucleic acids and proteins. Microfluid Nanofluid 10(2):231–247

Choi JR, Hu J, Tang R, Gong Y, Feng S, Ren H, Wen T, Li X, Abas WABW, Pingguan-Murphy B, Xu F (2016) An integrated paper-based sample-to-answer biosensor for nucleic acid testing at the point of care. Lab Chip 16(3):611–621

Cordray MS, Richards-Kortum RR (2012) Emerging nucleic acid–based tests for point-of-care detection of malaria. The American journal of tropical medicine and hygiene 87(2):223

Czilwik G, Messinger T, Strohmeier O, Wadle S, Von Stetten F, Paust N, Roth G, Zengerle R, Saarinen P, Niittymäki J, McAllister K (2015) Rapid and fully automated bacterial pathogen

detection on a centrifugal-microfluidic LabDisk using highly sensitive nested PCR with integrated sample preparation. Lab Chip 15(18):3749–3759
Ding X, Yan Y, Li S, Zhang Y, Cheng W, Cheng Q, Ding S (2015) Surface plasmon resonance biosensor for highly sensitive detection of microRNA based on DNA super-sandwich assemblies and streptavidin signal amplification. Anal Chim Acta 874:59–65
Flatland B, Freeman KP, Vap LM, Harr KE (2013) ASVCP guidelines: quality assurance for point-of-care testing in veterinary medicine. Vet Clin Pathol 42(4):405–423
Gandelman OA, Church VL, Moore CA, Kiddle G, Carne CA, Parmar S, Jalal H, Tisi LC, Murray JA (2010) Novel bioluminescent quantitative detection of nucleic acid amplification in real-time. PLoS One 5(11):e14155
Gill P, Ghaemi A (2008) Nucleic acid isothermal amplification technologies—a review. Nucleosides Nucleotides Nucleic Acids 27(3):224–243
Gootenberg JS, Abudayyeh OO, Kellner MJ, Joung J, Collins JJ, Zhang F (2018) Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. Science 360(6387): 439–444
Gopinath SCB (2007) Antiviral aptamers. Arch Virol 152:2137–2157
Gotham D, McKenna L, Deborggraeve S, Madoori S, Branigan D (2021) Public investments in the development of GeneXpert molecular diagnostic technology. PLoS One 16(8):e0256883
Haleyur Giri Setty MK, Hewlett IK (2014) Point of care technologies for HIV. AIDS Res Ther 2014:1
Han C, Li Q, Ji H, Xing W, Zhang L, Zhang L (2021) Aptamers: the powerful molecular tools for virus detection. Chem Asian J 16(11):1298–1306
Hansen G, Marino J, Wang ZX, Beavis KG, Rodrigo J, Labog K, Westblade LF, Jin R, Love N, Ding K, Garg S (2021) Clinical performance of the point-of-care cobasLiat for detection of SARS-CoV-2 in 20 minutes: a multicenter study. J Clin Microbiol 59(2):e02811–e02820
Harrington LB, Burstein D, Chen JS, Paez-Espino D, Ma E, Witte IP, Cofsky JC, Kyrpides NC, Banfield JF, Doudna JA (2018) Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. Science 362(6416):839–842
Hsieh K, Melendez JH, Gaydos CA, Wang TH (2022) Bridging the gap between development of point-of-care nucleic acid testing and patient care for sexually transmitted infections. Lab Chip 22:476
Hu J, Wang S, Wang L, Li F, Pingguan-Murphy B, Lu TJ, Xu F (2014) Advances in paper-based point-of-care diagnostics. Biosens Bioelectron 54:585–597
Huang Y, Zhang L, Zhang H, Li Y, Liu L, Chen Y, Qiu X, Yu D (2020) Development of a portable SPR sensor for nucleic acid detection. Micromachines 11(5):526
Jangam SR, Agarwal AK, Sur K, Kelso DM (2013) A point-of-care PCR test for HIV-1 detection in resource-limited settings. Biosens Bioelectron 42:69–75
Jin J, Duan L, Fu J, Chai F, Zhou Q, Wang Y, Shao X, Wang L, Yan M, Su X, Zhang Y (2021) A real-time LAMP-based dual-sample microfluidic chip for rapid and simultaneous detection of multiple waterborne pathogenic bacteria from coastal waters. Anal Methods 13(24):2710–2721
Joshi VG, Chindera K, Singh AK, Sahoo AP, Dighe VD, Thakuria D, Tiwari AK, Kumar S (2013) Rapid label-free visual assay for the detection and quantification of viral RNA using peptide nucleic acid (PNA) and gold nanoparticles (AuNPs). Anal Chim Acta 795:1–7
Kaur N, Toley BJ (2018) Based nucleic acid amplification tests for point-of-care diagnostics. Analyst 143(10):2213–2234
Kim SA, Kim SJ, Lee SH, Park TH, Byun KM, Kim SG, Shuler ML (2009) Avian influenza-DNA hybridization detection using wavelength interrogation-based surface plasmon resonance biosensor. In: Sensors, 2009 IEEE. IEEE, pp 832–835
Kim YS, Chung J, Song MY, Jurng J, Kim BC (2014) Aptamer cocktails: enhancement of sensing signals compared to single use of aptamers for detection of bacteria. Biosens Bioelectron 54: 195–198
Kim YT, Lee D, Heo HY, Sim JE, Woo KM, Im SG, Seo TS (2016) Total integrated slidable and valveless solid phase extraction-polymerase chain reaction-capillary electrophoresis

microdevice for mini Y chromosome short tandem repeat genotyping. Biosens Bioelectron 78: 489–496
Kocak DD, Gersbach CA (2018) From CRISPR scissors to virus sensors, vol 557, p 168
Konermann S, Lotfy P, Brideau NJ, Oki J, Shokhirev MN, Hsu PD (2018) Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. Cell 173(3):665–676
Koteswara Rao V (2021) Point of care diagnostic devices for rapid detection of novel coronavirus (SARS-nCoV19) pandemic: a review. Front Nanosci 2:593619
Kumar N, Bhatia S, Pateriya AK, Sood R, Nagarajan S, Murugkar HV, Kumar S, Singh P, Singh VP (2020) Label-free peptide nucleic acid biosensor for visual detection of multiple strains of influenza a virus suitable for field applications. Anal Chim Acta 1093:123–130
Labib M, Zamay AS, Kolovskaya OS, Reshetneva IT, Zamay GS, Kibbee RJ, Sattar SA, Zamay TN, Berezovski MV (2012) Aptamer-based viability impedimetric sensor for bacteria. Anal Chem 84(21):8966–8969
Lee HH, Dineva MA, Chua YL, Ritchie A, Ushiro-Lumb I, Wisniewski CA (2010) Simple amplification-based assay: a nucleic acid-based point-of-care platform for HIV-1 testing. J Infect Dis 201:S65–S71
Li J, Macdonald J (2015) Advances in isothermal amplification: novel strategies inspired by biological processes. Biosens Bioelectron 64:196–211
Lim GS, Chang JS, Lei Z, Wu R, Wang Z, Cui K, Wong S (2015) A lab-on-a-chip system integrating tissue sample preparation and multiplex RT-qPCR for gene expression analysis in point-of-care hepatotoxicity assessment. Lab Chip 15(20):4032–4043
Lin CW, Lee JH, Chiu NF, Lee SY, Liu KC, Tsai FY, Yen CY, Lee CN (2009) Hybrid nano plasmonics for integrated biosensor. In: 2009 Asia Communications and Photonics conference and Exhibition (ACP), vol 2009. IEEE, pp 1–6
Linnes JC, Fan A, Rodriguez NM, Lemieux B, Kong H, Klapperich CM (2014) Based molecular diagnostic for chlamydia trachomatis. RSC Adv 4(80):42245–42251
Lu W, Chen Y, Liu Z, Tang W, Feng Q, Sun J, Jiang X (2016) Quantitative detection of microRNA in one step via next generation magnetic relaxation switch sensing. ACS Nano 10(7):6685–6692
Mautner L, Baillie CK, Herold HM, Volkwein W, Guertler P, Eberle U, Ackermann N, Sing A, Pavlovic M, Goerlich O, Busch U (2020) Rapid point-of-care detection of SARS-CoV-2 using reverse transcription loop-mediated isothermal amplification (RT-LAMP). Virol J 17(1):1–14
Misono TS, Kumar PK (2005) Selection of RNA aptamers against human influenza virus hemagglutinin using surface plasmon resonance. Anal Biochem 342(2):312–317
Miyamoto S, Sano S, Takahashi K, Jikihara T (2015) Method for colorimetric detection of double-stranded nucleic acid using leuco triphenylmethane dyes. Anal Biochem 473:28–33
Murtaza G, Rizvi AS, Xue M, Qiu L, Meng Z (2022) Consensus receptor-binding domain-targeted aptamer selection and designing of a photonic crystal-decorated Aptasensor for SARS-CoV-2. Anal Biochem 94:7391
Myhrvold C, Freije CA, Gootenberg JS, Abudayyeh OO, Metsky HC, Durbin AF, Kellner MJ, Tan AL, Paul LM, Parham LA, Garcia KF (2018) Field-deployable viral diagnostics using CRISPR-Cas13. Science 360(6387):444–448
Najian AN, Syafirah EEN, Ismail N, Mohamed M, Yean CY (2016) Development of multiplex loop mediated isothermal amplification (m-LAMP) label-based gold nanoparticles lateral flow dip-stick biosensor for detection of pathogenic Leptospira. Anal Chim Acta 903:142–148
Nayak S, Blumenfeld NR, Laksanasopin T, Sia SK (2017) Point-of-care diagnostics: recent developments in a connected age. Anal Chem 89(1):102–123
Niemz A, Ferguson TM, Boyle DS (2011) Point-of-care nucleic acid testing for infectious diseases. Trends Biotechnol 29(5):240–250
Notomi T, Mori Y, Tomita N, Kanda H (2015) Loop-mediated isothermal amplification (LAMP): principle, features, and future prospects. J Microbiol 53(1):1–5
Nsamela A (2020) Microfluidics for PLGA nanoparticle synthesis: a review. Elveflow

Oh SY, Heo NS, Shukla S, Cho HJ, Vilian AT, Kim J, Lee SY, Han YK, Yoo SM, Huh YS (2017) Development of gold nanoparticle-aptamer-based LSPR sensing chips for the rapid detection of Salmonella typhimurium in pork meat. Sci Rep 7(1):1–10

Palau W, Masante C, Ventura M, Di Primo C (2013) Direct evidence for RNA–RNA interactions at the 3′ end of the hepatitis C virus genome using surface plasmon resonance. RNA 19(7): 982–991

Pardee K, Green AA, Takahashi MK, Braff D, Lambert G, Lee JW, Ferrante T, Ma D, Donghia N, Fan M, Daringer NM (2016) Rapid, low-cost detection of Zika virus using programmable biomolecular components. Cell 165(5):1255–1266

Peeling RW, Holmes KK, Mabey D, Ronald A (2006) Rapid tests for sexually transmitted infections (STIs): the way forward. Sex Transm Infect 82:v1–v6

Petralia S, Sciuto EL, Conoci S (2017) A novel miniaturized biofilter based on silicon micropillars for nucleic acid extraction. Analyst 142(1):140–146

Rezk AR, Qi A, Friend JR, Li WH, Yeo LY (2012) Uniform mixing in paper-based microfluidic systems using surface acoustic waves. Lab Chip 12(4):773–779

Rodriguez NM, Linnes JC, Fan A, Ellenson CK, Pollock NR, Klapperich CM (2015) Based RNA extraction, in situ isothermal amplification, and lateral flow detection for low-cost, rapid diagnosis of influenza a (H1N1) from clinical specimens. Anal Chem 87(15):7872–7879

Rohrman BA, Richards-Kortum RR (2012) A paper and plastic device for performing recombinase polymerase amplification of HIV DNA. Lab Chip 12(17):3082–3088

Saini K, Kaushal A, Gupta S, Kumar D (2019) Rapid detection of Salmonella enterica in raw milk samples using Stn gene-based biosensor. *3 Biotech* 9(11):1–9

Schito M, Migliori GB, Fletcher HA, McNerney R, Centis R, D'Ambrosio L, Bates M, Kibiki G, Kapata N, Corrah T, Bomanji J (2015) Perspectives on advances in tuberculosis diagnostics, drugs, and vaccines. Clin Infect Dis 61:S102–S118

Shafiei F, McAuliffe K, Bagheri Y, Sun Z, Yu Q, Wu R, You M (2020) Based fluorogenic RNA aptamer sensors for label-free detection of small molecules. Anal Methods 12(21):2674–2681

Shmakov S, Abudayyeh OO, Makarova KS, Wolf YI, Gootenberg JS, Semenova E, Minakhin L, Joung J, Konermann S, Severinov K, Zhang F (2015) Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. Mol Cell 60(3):385–397

Song J, Mauk MG, Hackett BA, Cherry S, Bau HH, Liu C (2016) Instrument-free point-of-care molecular detection of Zika virus. Anal Chem 88(14):7289–7294

Su X, Kanjanawarut R (2009) Control of metal nanoparticles aggregation and dispersion by PNA and PNA− DNA complexes, and its application for colorimetric DNA detection. ACS Nano 3(9):2751–2759

Tang XY, Zhang HY (2006) Advances in the application of SPR *Biosens*. China Anim Husb Vet Med 33:47–50

Tyas AA, Raeni SF, Sakti SP, Sabarudin A (2021) Recent advances of hepatitis B detection towards paper-based analytical devices. Sci World J 2021:6643573

Vashist SK (2017) Point-of-care diagnostics: recent advances and trends. Biosensors 7(4):62

Wang R, Zhao X, Chen X, Qiu X, Qing G, Zhang H, Zhang L, Hu X, He Z, Zhong D, Wang Y (2019) Rolling circular amplification (RCA)-assisted CRISPR/Cas9 cleavage (RACE) for highly specific detection of multiple extracellular vesicle MicroRNAs. Anal Chem 92(2): 2176–2185

Wang Y, Qi W, Wang L, Lin J, Liu Y (2021) Magnetic bead chain-based continuous-flow DNA extraction for microfluidic PCR detection of salmonella. Micromachines 12(4):384

Websource (2016) p. 108. Report ID: 978–1–68038-725-4. http://www.grandviewresearch.com/press-release/global-paper-diagnostics-market

Wong YP, Othman S, Lau YL, Radu S, Chee HY (2018) Loop-mediated isothermal amplification (LAMP): a versatile technique for detection of micro-organisms. J Appl Microbiol 124(3): 626–643

Xu G, Hsieh TM, Lee DY, Ali EM, Xie H, Looi XL, Koay ESC, Li MH, Ying JY (2010) A self-contained all-in-one cartridge for sample preparation and real-time PCR in rapid influenza diagnosis. Lab Chip 10(22):3103–3111

Yan L et al (2014) Isothermal amplified detection of DNA and RNA. Mol BioSyst 10:970–1003

Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, Van Der Oost J, Regev A, Koonin EV (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163(3):759–771

Zhang L, Zhang Y, Wang C, Feng Q, Fan F, Zhang G, Jiang X (2014) Integrated microcapillary for sample-to-answer nucleic acid pretreatment, amplification, and detection. Anal Chem 86(20): 10461–10466

Part III
Bioinformatics, Big Data, and Integrated Omics

Chapter 16
Bioinformatics: Unveiling the Systems Biology

Amit Kumar and Sheikh Firdous Ahmad

Abstract The human population is continuously increasing, and food production has to keep pace ensuring food and nutritional security. Genetics and biotechnology fields have played a significant role in improving food production by exploiting the variation existent in the agricultural and livestock population. Traditional breeding approaches have the inherent limitation of increasing generation intervals and correspondingly lesser genetic gains. The penetration of the latest data-based technologies and access to the latest bioinformatics and statistical tools have truly revolutionized biological processes via the analysis of nucleic acid molecules like DNA, RNA, and protein. Multiple omics-based technologies have deeply penetrated the current scientific arena. These include genomics, transcriptomics, proteomics, and metabolomics. An amalgamation of biology, computational science, bioinformatics, and other related fields is desired to gain maximum insights into biological mechanisms with increased certainty. Systems biology provides one such avenue that involves the application of mathematical and computational modeling to solve biological problems and understand biological processes through an interdisciplinary approach. The field of systems biology finds its applications in various aspects of modern science varying from deducing genetic diversity parameters to elucidating the potential association of genetic variants with traits of economic interest or pathophysiological states including diseases. The present chapter aims to introduce the concepts of bioinformatics, and systems biology and then discuss the details of various Omics-based technologies in detail.

Keywords Bioinformatics · Systems biology · Omics research · Molecular breeding

A. Kumar (✉) · S. F. Ahmad
Division of Animal Genetics, ICAR-Indian Veterinary Research Institute, Bareilly, Uttar Pradesh, India
e-mail: sheikh.ahmad@icar.gov.in

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_16

16.1 Introduction

Human civilization is ever-increasing with a major contribution from developing countries like China and India. Increased population poses certain challenges to its existence with noticeable threats erupting from changing agro-climatic trends and food habits (Tuomisto et al. 2017). The existence of human civilization is hugely dependent on agriculture and allied sectors. Agriculture and allied sectors support human civilization by ensuring food and nutritional security for the masses (Swaminathan and Bhavani 2013). Human civilization has progressed in various forms and improved food production from agriculture and allied activities by using different scientific technologies.

Variation forms an important part that is exploited for improvement in food production via techniques from genetics and biotechnology fields to keep pace with increased food demand. The progress of human civilization in terms of adopting better practices related to agriculture, livestock, and allied sectors has been tremendous. Human civilization has directed most of its efforts toward exploiting the variation existing in the universe, be it milk productivity of dairy animal populations, fruit production of horticulture crops, cereal production of agricultural fields, or value addition to agricultural produce. Traditional breeding strategies have mainly been used to exploit the underlying variation in various crop and livestock populations.

Traditional breeding approaches suffer from the significant hindrance of being time-consuming which contributes to increased generation interval in crops and livestock populations. Gradually, molecular markers entered their way into the arena and helped in understanding and exploiting qualitative traits in a better way. However, studying quantitative traits using molecular markers is difficult as they are affected by multiple genes as well as the environment to a considerable level (Anderson et al. 2013). The individual effect of these genes is too small to be captured accurately (Nakaya and Isobe 2012), although the combined effect is large enough to be quantified. Marker-assisted selection (MAS) entails the use of molecular markers for selecting/ culling individuals at an early age (Ashraf et al. 2012). The field of genetics and biotechnology has seen a tremendous transformation with the advent and advancement of molecular markers. The transformation started with the discovery of random fragment length polymorphism (RFLP) as a first-generation molecular marker. Improved genetic gains have been realized recently with the advent of efficient and computationally amenable molecular markers coupled with reliable detection of their association with QTLs affecting traits of economic interest. Recently, microsatellite and single nucleotide polymorphism (SNP) markers have been increasingly used for various purposes in genetics and breeding. Microsatellites are multi-loci markers that are 2–6 bp tandem repeats and are expressed in a co-dominant fashion (Abdul-Muneer 2014; Vieira et al. 2016). Microsatellite markers have been recommended by the Food and Agricultural Organization (FAO) for genetic diversity studies in various species of food animals (Khade et al. 2019). On the other hand, SNPs are single-base substitutions wherein

one nucleotide may have different polymorphic forms at the population level. In a single position, four polymorphic forms are possible for a particular SNP; however, two forms occur practically in most situations. Millions of SNPs have been reported to occur in different livestock species (Wilkinson et al. 2011). These molecular markers are informative with high polymorphism information content (PIC), present near-uniformly throughout the genome, and are amenable to bioinformatics and statistical tools (Schaid et al. 2004; Tokarska et al. 2009; Ling et al. 2020;).

Courtesy of the advent of molecular markers, scientific research has progressed from studying genes, genetic variants, and proteins individually with consequent effect on the phenotype to analyzing the whole genome and proteome with considerable accuracy, reliability, and reproducibility. The immensely paced generation of high-throughput data in the current era has truly revolutionized the understanding of biological processes via the analysis of nucleic acid molecules like DNA, RNA, and protein. It has revolutionized how cells and cellular processes are studied. Multiple omics-based technologies have deeply penetrated the current scientific arena. These include genomics, transcriptomics, proteomics, and metabolomics (Wu et al. 2022). It also includes the application and amalgamation of different branches of science like immunogenetics and immunogenomics; nutrigenetics and nutrigenomics; pharmacogenomics, and other similar branches. Several statistical and bioinformatics tools are now available to study the same biological phenomenon *vis-à-vis* different fields of science.

Bioinformatics refers to the discipline of science that deals with the application of computational tools and statistics to solve biological problems. It entails the processes of acquisition storage, analysis, and dissemination of data about biological problems (Bayat 2002). The field of bioinformatics has progressed immensely during the last few decades, providing much-needed support to the introduction and propagation of "omics"-based technologies. Deep insights have been deduced based on big data that is generated on modern platforms. Big data analytics *vis-à-vis* biological systems in humans and animals need immense support from computational science. It is an amalgamation of biology, computational science, bioinformatics, and other related fields that are destined to help gain maximum insights into biological mechanisms with increased certainty. It may also help in understanding various aspects of biology for overall human and animal welfare. Systems biology provides one such avenue wherein the intersection of these fields is possible. The field of systems biology has recently emerged in the era of cross-disciplinary biological research. Systems biology refers to the application of mathematical and computational modeling to solve biological problems and understand biological processes through an interdisciplinary approach (Likić et al. 2010). With a plethora of data generation opportunities available, it is indispensable to integrate the information from all aspects that may help to gain deep insights into the control and regulation of biological processes in body systems. The systems biology approach possesses the potential to help reap maximum gains in this direction.

The field of systems biology finds its applications in various aspects of modern science varying from deducing genetic diversity parameters to elucidating the potential association of genetic variants with traits of economic interest or

pathophysiological states including diseases. These applications are based on various approaches to studying diverse aspects of the biological makeup of an organism at different levels of central dogma, i.e., replication (DNA), transcription (RNA), and translation (protein). Therefore, the modern approaches to systems biology mainly include studies on genomics, transcriptomics, and proteomics (Pinu et al. 2019).

16.2 Genomics, Bioinformatics, and Systems Biology

Genome refers to the haploid set of chromosomes with gene content within an individual (Little 2005). It contains both chromosomal and extra-chromosomal DNA, whether it is coding or non-coding in nature. An individual's genome acts as an information repository that encodes for all its cellular and developmental life processes (Goldman and Landweber 2016). Exploration of individuals' genetic makeup at various levels while using techniques from different scientific fields including bioinformatics has helped unravel complex biological phenomena. Genomics refers to the field of science wherein dense SNP genotype data at the genome scale are used to gain maximum insights into different biological processes or elucidate the genetic variants responsible for variation in the specific trait of interest (Ombrello et al. 2014). The completion of the Human Genome Project in 2003 and the preparation of the human draft genome have produced a huge impact on our understanding of the interaction and synergistic effects of the genome to produce a definite phenotype (Khodadadian et al. 2020). Genomic selection refers to the process of selecting/ rejecting an individual based on its genetic makeup in terms of dense marker data generated on a genome-wide scale (Meuwissen et al. 2001). It works on the principle that at least one of the dense markers will be in linkage disequilibrium (LD) with the quantitative trait loci (QTLs) that affect the trait of interest (Goddard and Hayes 2009). Linkage disequilibrium refers to the non-random association between alleles at different loci within an individual (Kang and Rosenberg 2019). The field of genomics has progressed manifolds with genome-wide SNP markers being used for various purposes including elucidating the genetic diversity of a population, population structure, deducing copy number variation map, haplotype structure, elucidating selection sweeps (within and across the population), and ultimately delineating association between SNP and phenotype of interest. With genome-wide data available for an individual, it is possible to predict, with considerable accuracy, its phenotype that is likely to be expressed at later life stages. Accordingly, the informed decision may be made at an early stage without having to wait for the life stage and phenotype to appear. The studies on SNPs at the genome-wide level have made significant progress in the last decade, especially in developing countries due to the increased and easy availability of bioinformatics and statistical tools.

In recent times, genomic technologies have penetrated the field of animal breeding by promising significant improvements. These technologies have transformed the field of animal breeding by increasing the prediction accuracy up to the level of

0.8 and more for different traits (Lenz et al. 2017). It has opened up new facets and provides avenues for in-depth studies related to most economic traits of interest in major farm animal species. Implementation of genomic selection in livestock relies mainly on single nucleotide polymorphism (SNP) arrays. Currently, SNP arrays are commercially available for various farm animal species across different densities. The genomics field has progressed and is divided into three sub-fields, i.e., structural genomics, functional genomics, and comparative genomics.

16.2.1 Structural Genomics

Recently, the field of genomics has found many applications, especially in livestock genomics, for overall human and animal welfare. Structural genomics is a naive field of biology that refers to the exploration of the relationship between genetic bases, protein structures, and phenotype(s) at a genome-wide scale (Goldsmith-Fischman and Honig 2003). It involves the application of various experimental and computational approaches, exclusively or in combination. Information on gene sequence(s), protein structures, and phenotypic information help in deducing its biological bases. Furthermore, it provides the bases for a deep understanding of various biological phenomena involved in generating traits of varied nature. Structural genomics mainly involves two main activities, i.e., the construction of high-resolution maps based on data on molecular markers across the genome of a species and sequencing of the order of nucleotides at different levels.

Two approaches are mainly available to generate genome-wide data on molecular markers from nucleic acid material, i.e., genotyping and sequencing. Genotyping refers to the elucidation of genetic variants at fixed positions across the genome. The fixed positions are decided based on the occurrence of molecular markers in breeds/ populations used for the data generation. The data generated on these molecular markers are used to develop arrays or BeadChips wherein probes are fixed along with the chip. This approach, thus, needs prior knowledge about the draft genome and the location of molecular markers along with it. Subsequently, the DNA of an unknown individual can be used on the BeadChip to generate original genotyping data for that individual. On the other hand, sequencing refers to the elucidation of individual nucleotides and their order along a genetic sequence or complete gene. The genotyping approach helps investigate variation along the fixed segments/ portion within a sequence/ gene while the sequencing approach investigates each nucleotide along the segment. Sequencing can be done on a whole-genome basis by investigating every nucleotide present within the genome. This approach is termed whole-genome sequencing (WGS). However, it proves to be economically intensive and demands huge computational support for any interference about systems biology. Under an innovative sequencing technology that has been developed, single or two restriction enzymes are used to digest genomic DNA into fragments that are followed by their sequencing. This approach is commonly known as restriction site-associated DNA sequencing (RADseq). Genotyping and sequencing are the two

main aspects of structural genomics which act as a starting material to build upon and gain maximum insights into how the genetic makeup of an individual is responsible for variation at the phenotype level.

16.2.1.1 BeadChips for Livestock Species

The advent of whole-genome sequencing (WGS) and the availability of draft genomes for various farm animal species have enabled the identification of a large number of molecular markers including microsatellites, SNPs, and other Indel variations (Green 2001). It has, in turn, enabled researchers to use this genotype information for gaining maximum insights into biological processes. The availability of draft genomes for human and various farm animal species has revolutionized the studies related to GWAS, genomic selection, and other aspects. The Bovine HapMap Genome Project emerged as an effective launch pad for applications of genomics in livestock to characterize and disentangle the complexity of the bovine genome. The term "SNP BeadChip" refers to the collection of informative and dense SNPs in an array form that is fitted with probes that are specific for particular SNPs and can be used to generate information at these specific points in various livestock species. BeadChips are used to assess variation in terms of molecular markers present at specific locations of the genome. SNP BeadChips of various densities have been developed and are available commercially for different farm animal species. The SNPs used in these BeadChips have been identified through various techniques including novel SNPs generated via sequencing, Bovine HapMap dataset, *Btau* Assembly SNPs, whole-genome shotgun reads, and Holstein BAC sequence data among others. Illumina^{R}, Affytermix^{R}, and GeneSeek^{R} are noteworthy commercial enterprises that facilitate the availability and use of SNP BeadChip platforms all over the globe for the generation of genome-wide data. These SNP BeadChips have been developed and validated on different livestock breeds across the globe. These chips represent several thousand and millions of SNP markers and data are generated by using the principle of hybridization between probes and sample DNA.

Two platforms are routinely used for genotyping DNA samples from humans and farm animals, viz. Illumina (San Diego, CA) and Affymetrix (Santa Clara, CA). However, the major issue with the development and use of these BeadChips includes ascertainment bias that gets generated when chips developed on breeds of one lineage are used on breeds of another lineage (Ahmad et al. 2020). Bias is generated due to less representation of major lineages of a particular species during the identification of molecular markers at chip development stages. SNP BeadChips of varying densities is available now for different farm animal species (Table 16.1).

16.2.1.2 SNP Genotyping

SNP genotyping refers to the process of elucidation of the variation at specific locations between members of a population/ species when compared with some

Table 16.1 Availability of different BeadChips with corresponding densities in different livestock species

S. No.	Species	Available BeadChip	No. of markers (approximate)
1	Bovine	BovineSNP3K	~3000
		BovineSNP50K	54,001
		BovineHD777K	777,962
2	Bubaline	BubalineSNP90K	90,000
3	Ovine	OvineSNP50K	54,241
4	Caprine	CaprineSNP50K	53,347
5	Swine	PorcineSNP60K	64,332
6	Equine	EquineSNP50K	54,602
7	Dogs	CanineSNP20K	22,362
		CanineHD	170,000

Adopted from Lashmar et al. (2019) https://doi.org/10.4314/sajas.v49i2.7 and Fan et al. (2010) https://doi.org/10.5713/ajas.2010.r.03

standard. BeadChip is a micro-electro-mechanical system containing wells with probes attached to a large number of beads on a silicon wafer. The detection of variation in the query DNA sequence is based on the hybridization principle. The Infinium assay, which forms the basis of Illumina-based BeadChips, produces efficient results with considerable accuracy in a short period (Adler et al. 2013). DNA samples are initially isolated from various types of tissues that are then amplified via thermal cycling followed by their fragmentation. The DNA fragments hybridize against corresponding probes during the next 16–24 h, the results of which are assessed on iScan platform. SNP genotypes are called based on the color reaction generated on hybridization, its detection, and analyses on imaging technology (Gunderson et al. 2005).

The coverage and throughput of DNA sequencing have made significant progress while the cost of genotyping individuals for high-density genome-wide markers has plummeted. Improved technologies have enabled concurrent genotyping of a large number of individuals in a single go. Availability and usage of improved technologies, better bioinformatics, and statistical tools have helped in a deeper understanding of the complex genotype–phenotype relationship. The advent of commercially available SNP BeadChips of various densities has further propelled the research using genome-wide data. However, genotyping a large number of individuals concurrently is still limited, especially in developing countries, mainly due to reasons of genotyping cost and ascertainment bias. The markers used for the development of these chips may not be polymorphic in all breeds/ populations of a particular species reared around the globe. Moreover, SNP arrays demand considerable financial investments, especially at the initial stages, and are not flexible enough *vis-à-vis* experimental design.

Genomic predictions, related to association, diversity, population structure, selection signatures, copy number variation, etc., are made by fitting different models on genome-wide data. However, the statistical power of these predictions is limited by the density of markers and sample size, especially in developing countries. Several

technologies have been developed which indirectly help increase the sample size with a similar resource investment by reducing the genotyping or sequencing costs. These include designing low-density panels and comparing their efficiency *vis-à-vis* imputation accuracy and admixture predictions. Scientific literature suggests that 3–12 thousand markers may be enough to impute the rest of the ungenotyped/ missing markers with considerable efficiency and accuracy (Hayes et al. 2012; Wellmann et al. 2013). However, the accuracy and success are dependent on various factors including the species involved, the status of linkage disequilibrium within the population, strategy for marker selection, algorithm used for imputation and makeup of reference dataset, etc. (Ma et al. 2013; Sargolzaei et al. 2014; Grossi et al. 2018).

16.2.1.3 Restriction-Associated DNA Sequencing (RADseq)

Genomic selection requires a large number of individuals to be genotyped or sequenced reliably and conveniently. It is difficult to genotype or sequence an individual across the whole of its genome (whole-genome selection) due to its cost-intensive nature; it also produces a large amount of data which poses additional bioinformatics and statistical challenges (Belkadi et al. 2015; Wang et al. 2018). Each individual being studied needs to be genotyped or sequenced for the same genetic elements for making meaningful inferences. One of the earliest techniques to ensure each individual is genotyped for the same regions across the genome was through the usage of restriction enzymes which led to the construction of "reduced representation" sequencing libraries (Torkamaneh et al. 2020). Approaches involving high-throughput and parallel sequencing soon replaced earlier technologies. Improved technologies involving second-generation sequencing approaches were based on libraries from genomic regions adjacent to restriction sites and were termed restriction-associated DNA sequencing (RADseq). Better methods of sequencing and genotyping have evolved which are based on the availability of complete reference sequence/ genome information. However, only fewer approaches including RADseq are currently available which are capable to work with both model and non-model species, irrespective of the presence of their reference genome.

Given the vastness of the genome, especially of humans and mammals, it is difficult to sequence each base individually. Therefore, a fraction of the genome is selected through different approaches for sequencing. However, it becomes challenging to ensure the uniformity of genomic regions being sequenced in all individuals of a population. One such optimal approach is provided by restriction site-associated DNA sequencing (RADseq) wherein the genomic regions adjacent to restriction enzyme cut sites are sequenced and analyzed (Peterson et al. 2012). It uses restriction enzyme digestion and sequencing of adjacent regions to examine the same subset of genomic regions and identify different types of polymorphism. Though other approaches of next-generation sequencing (NGS) examine larger sections of the genome and produce more data, however, they are cost-intensive and thus cannot be used for genotyping a large number of individuals.

ddRADseq is a variation of RAD sequencing protocol useful for SNP discovery and genotyping of individuals. It enables the combined usage of high-throughput genome-wide genotyping and next-generation sequencing technologies. Consequently, it has become increasingly popular in genome-wide analysis of livestock species, mainly due to their flexibility and relatively lower cost. As compared to the RADseq protocol, the fragment shearing step is replaced with second restriction digestion in ddRADseq to improve the accuracy of size selection (Wickland et al. 2017). A second index step is included to allow combinatorial indexing. Initially, genomic DNA is digested with restriction enzymes and barcoding of fragments with an adaptor is undertaken. Sequencing data of single-end reads are obtained from restriction sites, which is a great advantage in the sequencing of identical loci across multiple samples. However, using paired-end reads, the efficiency of mapping reads to the reference genome is increased (Donato et al. 2021). This gains significance, especially in species with complex genomes. Optionally, the fragments are combined if multiplexing of samples is needed. Subsequently, DNA is fragmented with a second restriction enzyme. This is followed by size selection and purification of fragments. Primers are used for the amplification of fragments as the next step. ddRADseq protocol helps reduce the cost and time to prepare the sequencing libraries. It also enables the sequencing of reads in paired-end mode and multiple sample analysis in concurrent mode. Therefore, ddRADseq has definite advantages over RADseq technology. Different *in silico* tools are currently available for the selection of restriction enzymes for ddRAD protocols.

16.2.1.4 Genome-Wide Association Studies (GWAS)

Genome-wide association studies (GWAS) aim to identify the differences in allele frequencies of single nucleotide polymorphism (SNPs) that vary at the genome-wide level and are systematically associated with a phenotypic trait of interest or complex traits (Marees et al. 2018). The availability of dense SNP chips of various densities in various species has propelled GWAS-related research in different species. The ultimate goal of GWAS is to identify those genetic regions that are responsible for the presence/absence of particular qualitative traits or a disease (Uffelmann et al. 2021). On the other hand, for quantitative traits, GWAS is destined to help identify those genetic regions that are responsible for increased or decreased levels of a specific phenotype. The phenotypic trait of interest used in GWAS may be quantitative or qualitative.

GWAS involves scanning the whole genome of an organism and identification of genetic markers, using different bioinformatics and statistical tools, so that outcome of a qualitative (such as the occurrence of a disease or not) or quantitative (milk production, etc.) trait. It involves the scanning of thousands and millions of markers (mostly SNPs) placed across the chromosomes present in respective species. The significance and confidence interval of such predictions hold pivotal importance. After the identification of genomic regions/ SNPs affecting the phenotype, the underlying biological pathways and networks can be elucidated efficiently.

Each SNP sequenced (without missing rate) is analyzed statistically to identify its potential association with the phenotype (Cano-Gamez and Trynka 2020). The phenotype should be properly defined and measurable on a definite scale. However, precautions should be undertaken during analysis as GWAS is susceptible to producing false-positive results. Therefore, adjustments may be needed before deducing conclusions and inferences from GWAS results. These include selecting an optimal statistical design, conducting multiple tests on the same datasets, adjusting for systematic stratifications due to genetic ancestry, adjusting for non-genetic factors, combining results from multiple GWAS studies, and using dense marker data along with accurate phenome information and finally validating the results using appropriate techniques.

Various software packages are currently available for genome-wide association studies in human and animal populations. The analysis for quantitative and qualitative traits varies in a few aspects. In Genome-wide Association studies, reliable and accurate phenome records are essential. The phenome data used for GWAS should cover the majority of variation existent in that trait within a particular population (Visscher et al. 2017). The more the variation covered in the phenomic records, the better will be the results from GWAS. After phenome recording, the genotyping is done to generate data on SNP or other suitable markers present within the genome of the organism. Finally, bioinformatics and statistical tools are used to associate phenome with SNP variations existent within the genome. The variation identified during GWAS should be consistent enough and not a false-positive result. Some of the software used for carrying out GWAS in human and livestock populations are PLINK (Purcell et al. 2007), TASSEL (Bradbury et al. 2007), and GAPIT (Lipka et al. 2012).

16.2.1.5 Copy Number Variation

Structural variations (SVs) have been found to play a significant role in creating differences among individuals of a population and cause variation at the phenotype level. They produce a significant effect on the evolution of a race. The size of structural variants varies from a few kilobases to a few megabases (Zhou et al. 2016). The existence of SVs is mainly attributable to events like deletion, duplication, translocation and inversion, and other similar events (Hou et al. 2011; Geistlinger et al. 2018). Copy number variation (CNV) is one of the prime structural variants that can cause disrupted or enhanced gene expression, ultimately affecting the phenotype. In other words, the CNVs affect the dosage of genes, expose recessive genes, and play a significant role in phenotypic diversity (Chen et al. 2012). Overlapping CNVs are routinely concatenated into CNV regions (CNVRs) and subsequently explored for their potential association with diverse phenotypes. The normal copy number of a gene or genetic variant is 2 while copy numbers of 0 and 1 are termed copy number loss and occur due to the deletion of gene/gene segments whereas the copy number above 2 (CN ~ 3 or 4) is termed copy number gain that occurs due to duplication of a gene(s)/ gene segments. CNVs at the

genome-wide level additionally possess evolutionary significance and affect the expression of genes in the vicinity besides functionally affecting the gene expression levels (Ahmad et al. 2022).

Karyotyping and fluorescent in-situ hybridization are important traditional techniques to elucidate the copy number variation of different genes at the cytogenetic level (Zhao et al. 2013). However, different bioinformatics software is now available to gain maximum insights into the copy number variations and copy number regions at the genome-wide level based on data from WGS or NGS, or array-based platforms. Various algorithms help in deducing the CNV and CNVR structure from genome-wide data generated across different platforms. These CNVs and CNVRs can be explored for association with traits of economic importance (Hou et al. 2012; Zhao et al. 2013). However, effective steps of detection, filtration, visualization, and annotation are needed to discover truly putative CNVs and CNVRs, irrespective of the algorithm used. The elucidation of CNVs and CNVRs from sequencing data is mainly dependent on the differential read depth of sequence fragments in the studied samples. Among all methods and algorithms, comparative genomic hybridization is considered the gold standard method of detecting CNVs and CNVRs (Geistlinger et al. 2018). On the other hand, the PennCNV program based on the hidden Markov model methodology has been recently developed and is considered one of the accurate algorithms for CNV and CNVR detection from array data, that too with improved sensitivity and specificity of detection (Zhang et al. 2014; Pierce et al. 2018).

16.2.2 Functional Genomics

The field of functional genomics entails the studies on structure, function, and regulation of genes at the genome level leading to the expression of phenotype (Auffray et al. 2003). It also entails the dynamic aspects of gene expression and regulation via transcription and translation processes along with interactions at various levels. The field of functional genomics aims to establish a relationship between the genome and the phenotype that is eventually expressed by an organism. Furthermore, functional genomics aims to understand how the genetic makeup (in terms of genes, intergenic regions, and other genetic variants along with interaction among them) at the genome level regulates various biological processes and eventually leads to varied phenotypic expression. Studying all genes/ genetic makeup of an organism concurrently is an important aspect of functional genomics with the hope of ultimately pinpointing a few putative candidate genes/ genetic variants that affect the phenotype, which can then be studied in further detail. It, thus, differs from the classical molecular biology approach wherein genes are studied individually. Functional genomics can be studied at all steps of central dogma; it, therefore, involves comprehensive studies done at genomics, epigenomics, transcriptomics, proteomics, and other levels and aimed at establishing a relation between phenotype and genomic composition and interactions thereof. It aims to

integrate the information from multiple processes related to central dogma and model dynamic networks leading to regulated phenotypic expression. It mainly studies the regulation by relating data with biological processes like differential gene expression, differentiation, and progression of cells (Bunnik and Le Roch 2013).

The approach of functional genomics has been used to gain deep insights into diverse phenotypes including qualitative (such as disease incidence) and quantitative (such as milk production) traits. Mostly, genome-wide association studies (GWAS) are used to elucidate the polygenic architecture of different pathophysiological states in humans and animals. GWAS aims to identify the causal genetic variants that are significantly associated (via linkage disequilibrium) with the gene of interest and eventually affect the incidence of a particular phenotype.

Techniques like microarray and next-generation sequencing (NGS) have produced a significant impact on the understanding of biological processes in the present era. The NGS technique is considered one of the biggest milestones that have signified the relevance of biotechnology and bioinformatics in modern science. These techniques are useful to study the differential gene expression in contrasting phenotypes. A microarray consists of thousands of tiny microscopic spots in form of predetermined sequences or probes that are immobilized on a solid surface for hybridization (Koltai and Weingarten-Baror 2008). The results are based on the color intensity generated by the hybridization of DNA fragments with immobilized probes (Bunnik and Le Roch 2013). However, one of the limitations of microarray is a saturation of intensity signals or limited fluorescence detection accuracy whereas NGS technologies are based on fragmenting the nucleic acid material into a fragment and then sequencing the smaller fragments. The information generated thereof is used in gaining insights into the differential gene expression in phenotypes of contrast by comparing the read depth and transcript abundances. The intermediate steps of indexing and assembly are required during detailed NGS analyses. Various bioinformatics software is available during NGS analyses that help trim the poor-quality reads, index and assemble the transcriptome, and ultimately generate parameters linked to transcript abundances. Ultimately, the global transcript profile (expression profile) is generated that is associated with varied phenotypes. Upon normalization for different factors, the differentially expressed genes and genetic variants between contrasting phenotypes are determined. Other techniques used in functional genomics include RNA interference, mutagenesis, and genome annotation among others. One other field of genomics, known as comparative genomics, involves the understanding of complex genomes by applying information from non-model organisms.

16.2.2.1 Transcriptomics

Transcriptome refers to the collection of various kinds of RNAs within a specific cell or tissue (Lowe et al. 2017). The transcriptome profile is highly dynamic and is considered an efficient representation of the cellular state in terms of its

development, differentiation, regulatory processes, and diverse range of biological conditions (Khodadadian et al. 2020). In the current era, transcriptome profiling is within the access of the majority of researchers from different laboratories all over the globe. Transcriptome-based studies have become an integral part of studies aimed to deduce deep insights into different biological processes leading to contrasting phenotypes. The transcriptome is composed of varied RNA types varying from coding to non-coding ones and long stretches to smaller fragment RNAs. These factors thus warrant specialized steps and precautions during library preparation and bioinformatics-based analyses (Sun and Chen 2020). Studying transcriptome profiles under contrasting conditions helps understand the gene function, underlying regulatory pathways, and processes involved (Bunnik and Le Roch 2013). The approach of transcriptomics has been successfully used in understanding expression profiles during pathophysiological conditions in humans and animals, especially in cancer states. With an improved understanding of gene expression profiles, therapeutic interventions can be planned meticulously and effectively. This has, in fact, given rising to a new field of science, commonly known as precision medicine.

RNA-Seq refers to the collection of techniques that are aimed to study the gene expression and transcriptome profiles, assessing their abundances, and relating them to the contrasting phenotypes (Kukurba and Montgomery 2015). Nucleic acid material (RNA) is obtained from individuals with contrasting phenotypes which are fragmented into small pieces and fragments are sequenced individually. The fragment size is typically around 75–150 bp long; however, it can vary between 30 and 10,000 bp in length, depending upon the sequencing platform used. Various platforms are currently available for the generation of transcriptome data using the RNA-Seq procedure such as Illumina, SOLiD, Ion Torrent, PacBio, and NanoPore. These platforms vary in read length of transcripts produced, maximum throughput produced per run, and accuracy of sequencing (Liu et al. 2012; Quail et al. 2012). Unlike microarray technology, prior knowledge of the probes is not required in RNA-Seq data analyses. In other words, the sequences/ transcript variants assessed for their expression states in Microarray are predetermined while in the transcriptomics approach, all the sequences are assessed for their global expression.

The individual fragments are indexed and assembled using a reference genome/ transcriptome of the species. The number of reads mapping to a specific segment of the reference genome/ transcriptome is taken as its expression level. RNA-Seq data analyses involve quality control, alignment, quantification, and differential expression steps. In the current era, each step has its dedicated software(s) that are available under the open-access facility and provide a user-friendly interface. Novel genes, transcripts, and splice variants can also be elucidated using RNA-Seq data (Sharp 1994). It can also be helpful to analyze the gene boundaries and intron–exon junctions. The discovery of transcripts other than mRNA also helps in a better understanding of biological processes and their regulation in humans and animals. Besides, transcriptome profiling is applicable in disease and diagnostic profiling, studying pathogen response to different environments, precision medicine, gene function annotation, and studying non-coding RNAs.

16.2.2.2 Proteomics

Proteins from one of the important functional units of the cell are directly related to the expression of the phenotype. It is essential to understand the protein setup in terms of their generation, function, regulation, and interaction to understand biological processes better. The regulation of transcription and translation processes ensures that everything that is encoded at the DNA level is not expressed as phenotype at all times. Regulation of these processes is an important process to modulate the gene expression according to the situation and the needs. Therefore, the transcript levels are not proportionately correlated with protein abundances, and separate quantification at the translation level by assessment of protein levels gains significant importance (Liu et al. 2016). Besides, many proteins undergo post-translational modifications (Duan and Walther 2015), further complicating the structure and functioning of proteins and regulation of phenotype. The branch of science wherein the structure, functionality, and interaction of proteins are studied is referred to as proteomics. Gel electrophoresis and mass spectrophotometry along with their variants are considered indispensable for proteomics-based studies.

16.3 Conclusion

The high-throughput data technologies have penetrated deeply into the current scientific era. It is due to plummeting costs of generating big data along with increased access to bioinformatics and statistical tools, especially in developing countries. Systems biology entails the amalgamation of biology, computational science, bioinformatics, and other related fields and possesses the potential to help gain maximum insights into biological mechanisms through an interdisciplinary approach. Systems biology involves the studies of various aspects of central dogma involving various Omics-based technologies. Bioinformatics and systems biology have a significant role to play in the current and future scientific arena wherein complex biological problems will need to be solved and manipulations are done for improved human and animal betterment and welfare.

References

Abdul-Muneer PM (2014) Application of microsatellite markers in conservation genetics and fisheries management: recent advances in population structure analysis and conservation strategies. Genet Res Int 2014:1–11. https://doi.org/10.1155/2014/691759

Adler AJ, Wiley GB, Gaffney PM (2013) Infinium assay for large-scale SNP genotyping applications. J Vis Exp 81:e50683. https://doi.org/10.3791/50683

Ahmad SF, Panigrahi M, Ali A, Dar RR, Narayanan K, Bhushan B (2020) Evaluation of two bovine SNP genotyping arrays for breed clustering and stratification analysis in well-known taurine and

indicine breeds. Anim Biotechnol 31(3):268–275. https://doi.org/10.1080/10495398.2019.1578227

Ahmad SF, Singh A, Panda S, Malla WA, Kumar A, Dutt T (2022) Genome-wide elucidation of CNV regions and their association with production and reproduction traits in composite Vrindavani cattle. Gene 830:146510. https://doi.org/10.1016/J.GENE.2022.146510

Anderson JT, Wagner MR, Rushworth CA, Prasad KVSK, Mitchell-Olds T (2013) The evolution of quantitative traits in complex environments. Heredity 112(1):4–12. https://doi.org/10.1038/hdy.2013.33

Ashraf M, Akram NA, Mehboob-Ur-Rahman, Foolad MR (2012) Marker-assisted selection in plant breeding for salinity tolerance. Methods Mol Biol 913:305–333. https://doi.org/10.1007/978-1-61779-986-0_21

Auffray C, Imbeaud S, Roux-Rouquié M, Hood L (2003) From functional genomics to systems biology: concepts and practices. C R Biol 326(10–11):879–892. https://doi.org/10.1016/J.CRVI.2003.09.033

Bayat A (2002) Science, medicine, and the future: bioinformatics. BMJ 324(7344):1018. https://doi.org/10.1136/BMJ.324.7344.1018

Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, Shang L, Boisson B, Casanova JL, Abel L (2015) Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proc Natl Acad Sci U S A 112(17):5473–5478. https://doi.org/10.1073/PNAS.1418631112/SUPPL_FILE/PNAS.1418631112.SD02.XLSX

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23(19):2633–2635. https://doi.org/10.1093/BIOINFORMATICS/BTM308

Bunnik EM, Le Roch KG (2013) An introduction to functional genomics and systems biology. Adv Wound Care 2(9):490. https://doi.org/10.1089/WOUND.2012.0379

Cano-Gamez E, Trynka G (2020) From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. Front Genet 11:424. https://doi.org/10.3389/FGENE.2020.00424/BIBTEX

Chen C, Qiao R, Wei R, Guo Y, Ai H, Ma J, Ren J, Huang L (2012) A comprehensive survey of copy number variation in 18 diverse pig populations and identification of candidate copy number variable genes associated with complex traits. BMC Genomics 13(1):733. https://doi.org/10.1186/1471-2164-13-733

Donato L, Scimone C, Rinaldi C, D'Angelo R, Sidoti A (2021) New evaluation methods of read mapping by 17 aligners on simulated and empirical NGS data: an updated comparison of DNA- and RNA-Seq data from Illumina and ion torrent technologies. Neural Comput & Applic 33(22):15669–15692. https://doi.org/10.1007/S00521-021-06188-Z/TABLES/4

Duan G, Walther D (2015) The roles of post-translational modifications in the context of protein interaction networks. PLoS Comput Biol 11(2):1004049. https://doi.org/10.1371/JOURNAL.PCBI.1004049

Fan B, Du ZQ, Gorbach DM, Rothschild MF (2010) Development and application of high-density SNP arrays in genomic studies of domestic animals. Asian Australas J Anim Sci 23(7):833–847. https://doi.org/10.5713/AJAS.2010.R.03

Geistlinger L, Da Silva VH, Cesar ASM, Tizioto PC, Waldron L, Zimmer R, de Almeida Regitano LC, Coutinho LL (2018) Widespread modulation of gene expression by copy number variation in skeletal muscle. Sci Rep 8(1):1–11. https://doi.org/10.1038/s41598-018-19782-4

Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet 10(6):381–391. https://doi.org/10.1038/nrg2575

Goldman AD, Landweber LF (2016) What is a genome? PLoS Genet 12(7):e1006181. https://doi.org/10.1371/JOURNAL.PGEN.1006181

Goldsmith-Fischman S, Honig B (2003) Structural genomics: computational methods for structure analysis. Protein Sci 12(9):1813. https://doi.org/10.1110/PS.0242903

Green ED (2001) Strategies for the systematic sequencing of complex genomes. Nat Rev Genet 2:3. www.arabidopsis.org

Grossi DA, Brito LF, Jafarikia M, Schenkel FS, Feng Z (2018) Genotype imputation from various low-density SNP panels and its impact on accuracy of genomic breeding values in pigs. Animal 12(11):2235–2245. https://doi.org/10.1017/S175173111800085X

Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS (2005) A genome-wide scalable SNP genotyping assay using microarray technology. Nat Genet 37(5):549–554. https://doi.org/10.1038/ng1547

Hayes BJ, Bowman PJ, Daetwyler HD, Kijas JW, Van Der Werf JHJ (2012) Accuracy of genotype imputation in sheep breeds. Anim Genet 43(1):72–80. https://doi.org/10.1111/J.1365-2052.2011.02208.X

Hou Y, Liu GE, Bickhart DM, Cardone MF, Wang K, Kim ES, Matukumalli LK, Ventura M, Song J, VanRaden PM, Sonstegard TS, Van Tassell CP (2011) Genomic characteristics of cattle copy number variations. BMC Genomics 12(1):1–11. https://doi.org/10.1186/1471-2164-12-127

Hou Y, Bickhart DM, Hvinden ML, Li C, Song J, Boichard DA, Fritz S, Eggen A, DeNise S, Wiggans GR, Sonstegard TS, Van Tassell CP, Liu GE (2012) Fine mapping of copy number variations on two cattle genome assemblies using high density SNP array. BMC Genomics 13(1):1–10. https://doi.org/10.1186/1471-2164-13-376

Kang JTL, Rosenberg NA (2019) Mathematical properties of linkage disequilibrium statistics defined by normalization of the coefficient D = pAB − pApB. Hum Hered 84(3):127–143. https://doi.org/10.1159/000504171

Khade KA, Panigrahi M, Ahmad SF, Kumar P, Bhushan B (2019) Genetic characterization and assessment of diversity in Pandharpuri buffalo breed of India using heterologous microsatellite markers. Anim Biotechnol 31(5):426–431. https://doi.org/10.1080/10495398.2019.1612757

Khodadadian A, Darzi S, Haghi-Daredeh S, Eshaghi FS, Babakhanzadeh E, Mirabutalebi SH, Nazari M (2020) Genomics and transcriptomics: the powerful Technologies in Precision Medicine. Int J Gen Med 13:627. https://doi.org/10.2147/IJGM.S249970

Koltai H, Weingarten-Baror C (2008) Specificity of DNA microarray hybridization: characterization, effectors and approaches for data correction. Nucleic Acids Res 36(7):2395. https://doi.org/10.1093/NAR/GKN087

Kukurba KR, Montgomery SB (2015) RNA sequencing and analysis. Cold Spring Harb Protoc 2015(11):951. https://doi.org/10.1101/PDB.TOP084970

Lashmar SF, Muchadeyi FC, Visser C (2019) Genotype imputation as a cost-saving genomic strategy for south African Sanga cattle: A review. S Afr J Anim Sci 49(2):263–280. https://doi.org/10.4314/SAJAS.V49I2.7

Lenz PRN, Beaulieu J, Mansfield SD, Clément S, Desponts M, Bousquet J (2017) Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (*Picea mariana*). BMC Genomics 18(1):335. https://doi.org/10.1186/S12864-017-3715-5

Likić VA, McConville MJ, Lithgow T, Bacic A (2010) Systems biology: the next frontier for bioinformatics. Adv Bioinforma 2010:1–10. https://doi.org/10.1155/2010/268925

Ling C, Lixia W, Rong H, Fujun S, Wenping Z, Yao T, Yaohua Y, Bo Z, Liang Z (2020) Comparative analysis of microsatellite and SNP markers for parentage testing in the golden snub-nosed monkey (*Rhinopithecus roxellana*). Conserv Genet Resour 12(4):611–620. https://doi.org/10.1007/S12686-020-01147-7/TABLES/5

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28(18):2397–2399. https://doi.org/10.1093/BIOINFORMATICS/BTS444

Little PFR (2005) Structure and function of the human genome. Genome Res 15(12):1759–1766. https://doi.org/10.1101/GR.4560905

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M (2012) Comparison of next-generation sequencing systems. J Biomed Biotechnol 2012:1–11. https://doi.org/10.1155/2012/251364

Liu Y, Beyer A, Aebersold R (2016) On the dependency of cellular protein levels on mRNA abundance. Cell 165(3):535–550. https://doi.org/10.1016/J.CELL.2016.03.014

Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T (2017) Transcriptomics technologies. PLoS Comput Biol 13(5):e1005457. https://doi.org/10.1371/JOURNAL.PCBI.1005457

Ma P, Brøndum RF, Zhang Q, Lund MS, Su G (2013) Comparison of different methods for imputing genome-wide marker genotypes in Swedish and Finnish red cattle. J Dairy Sci 96(7):4666–4677. https://doi.org/10.3168/JDS.2012-6316

Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, Derks EM (2018) A tutorial on conducting genome-wide association studies: quality control and statistical analysis. Int J Methods Psychiatr Res 27(2):e1608. https://doi.org/10.1002/MPR.1608

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4):1819–1829. https://doi.org/10.1093/GENETICS/157.4.1819

Nakaya A, Isobe SN (2012) Will genomic selection be a practical method for plant breeding? Ann Bot 110(6):1303. https://doi.org/10.1093/AOB/MCS109

Ombrello MJ, Sikora KA, Kastner DL (2014) Genetics, genomics and their relevance to pathology and therapy. Best Pract Res Clin Rheumatol 28(2):175. https://doi.org/10.1016/J.BERH.2014.05.001

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for De novo SNP discovery and genotyping in model and non-model species. PLoS One 7(5):e37135. https://doi.org/10.1371/JOURNAL.PONE.0037135

Pierce MD, Dzama K, Muchadeyi FC (2018) Genetic diversity of seven cattle breeds inferred using copy number variations. Front Genet 9:163. https://doi.org/10.3389/fgene.2018.00163

Pinu FR, Beale DJ, Paten AM, Kouremenos K, Swarup S, Schirra HJ, Wishart D (2019) Systems biology and multi-omics integration: viewpoints from the metabolomics research community. Meta 9(4):76. https://doi.org/10.3390/METABO9040076

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, De Bakker PIW, Daly MJ, Sham PC (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81(3):559–575. https://doi.org/10.1086/519795

Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y (2012) A tale of three next-generation sequencing platforms: comparison of ion torrent, Pacific biosciences and Illumina MiSeq sequencers. BMC Genomics 13(1):1–13. https://doi.org/10.1186/1471-2164-13-341/FIGURES/5

Sargolzaei M, Chesnais JP, Schenkel FS (2014) A new approach for efficient genotype imputation using information from relatives. BMC Genomics 15(1):478. https://doi.org/10.1186/1471-2164-15-478

Schaid DJ, Guenther JC, Christensen GB, Hebbring S, Rosenow C, Hilker CA, McDonnell SK, Cunningham JM, Slager SL, Blute ML, Thibodeau SN (2004) Comparison of microsatellites versus single-nucleotide polymorphisms in a genome linkage screen for prostate cancer–susceptibility loci. Am J Hum Genet 75(6):948. https://doi.org/10.1086/425870

Sharp PA (1994) Split genes and RNA splicing. Cell 77(6):805–815. https://doi.org/10.1016/0092-8674(94)90130-9

Sun YM, Chen YQ (2020) Principles and innovative technologies for decrypting noncoding RNAs: from discovery and functional prediction to clinical application. J Hematol Oncol 13(1):1–27. https://doi.org/10.1186/S13045-020-00945-8/TABLES/3

Swaminathan MS, Bhavani RV (2013) Food production and availability - essential prerequisites for sustainable food security. Indian J Med Res 138(3):383. https://doi.org/10.1142/9789813200074_0025

Tokarska M, Marshall T, Kowalczyk R, Wójcik JM, Pertoldi C, Kristensen TN, Loeschcke V, Gregersen VR, Bendixen C (2009) Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: the case of European bison. Heredity 103(4):326–332. https://doi.org/10.1038/hdy.2009.73

Torkamaneh D, Boyle B, St-Cyr J, Légaré G, Pomerleau S, Belzile F (2020) NanoGBS: A miniaturized procedure for GBS library preparation. Front Genet 11:67. https://doi.org/10.3389/FGENE.2020.00067

Tuomisto HL, Scheelbeek PFD, Chalabi Z, Green R, Smith RD, Haines A, Dangour AD (2017) Effects of environmental change on agriculture, nutrition and health: A framework with a focus on fruits and vegetables. Wellcome Open Res 2:21. https://doi.org/10.12688/WELLCOMEOPENRES.11190.2

Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, Martin HC, Lappalainen T, Posthuma D (2021) Genome-wide association studies. Nat Rev Methods Primers 1(1):1–21. https://doi.org/10.1038/s43586-021-00056-9

Vieira MLC, Santini L, Diniz AL, Munhoz, C. de F. (2016) Microsatellite markers: what they mean and why they are so useful. Genet Mol Biol 39(3):312. https://doi.org/10.1590/1678-4685-GMB-2016-0027

Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J (2017) 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet 101(1):5. https://doi.org/10.1016/J.AJHG.2017.06.005

Wang Y, Li G, Ma M, He F, Song Z, Zhang W, Wu C (2018) GT-WGS: an efficient and economic tool for large-scale WGS analyses based on the AWS cloud service. BMC Genomics 19(1):89–98. https://doi.org/10.1186/S12864-017-4334-X/FIGURES/6

Wellmann R, Preuß S, Tholen E, Heinkel J, Wimmers K, Bennewitz J (2013) Genomic selection using low density marker panels with application to a sire line in pigs. Genet Sel Evol 45(1):28. https://doi.org/10.1186/1297-9686-45-28

Wickland DP, Battu G, Hudson KA, Diers BW, Hudson ME (2017) A comparison of genotyping-by-sequencing analysis methods on low-coverage crop datasets shows advantages of a new workflow, GB-eaSy. BMC Bioinformatics 18(1):586. https://doi.org/10.1186/S12859-017-2000-6

Wilkinson S, Wiener P, Archibald AL, Law A, Schnabel RD, McKay SD, Taylor JF, Ogden R (2011) Evaluation of approaches for identifying population informative markers from high-density SNP chips. BMC Genet 12:45. https://doi.org/10.1186/1471-2156-12-45

Wu L, Xie X, Liang T, Ma J, Yang L, Yang J, Li L, Xi Y, Li H, Zhang J, Chen X, Ding Y, Wu Q (2022) Integrated multi-omics for novel aging biomarkers and antiaging targets. Biomol Ther 12(1):39. https://doi.org/10.3390/BIOM12010039

Zhang X, Du R, Li S, Zhang F, Jin L, Wang H (2014) Evaluation of copy number variation detection for a SNP array platform. BMC Bioinformatics 15(1):1–9. https://doi.org/10.1186/1471-2105-15-50

Zhao M, Wang Q, Wang Q, Jia P, Zhao Z (2013) Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinformatics 14(SUPPL11):S1. https://doi.org/10.1186/1471-2105-14-S11-S1

Zhou Y, Utsunomiya YT, Xu L, Hay EHA, Bickhart DM, Sonstegard TS, Van Tassell CP, Garcia JF, Liu GE (2016) Comparative analyses across cattle genders and breeds reveal the pitfalls caused by false positive and lineage-differential copy number variations. Sci Rep 6(1):1–9. https://doi.org/10.1038/srep29219

Chapter 17
Bioinformatics in Development of Antivirals

Subhomoi Borkotoky and Debajit Dey

Abstract Virus infections and epidemics are becoming more common over the world, with new and old viruses resurfacing with fresh intensity. As a result of long development phases, allergic reactions, the emergence of resistant strains, and other factors, traditional drugs and vaccinations are ineffective in combating this threat. The use of bioinformatics has led to a reduction in overall risk and cost for antiviral drug discovery by reducing the time and effort put into various steps of drug discovery. In the wake of the world's largest epidemic of the twenty-first century, SARS-CoV-2, bioinformatics techniques have been applied in unprecedented ways encoding the molecular characteristics of infectious viruses, vaccine development, and drug discovery. Through remarkable advances in genome sequencing, protein expression, and high-throughput structure determination methods, massive amounts of data have become available to accelerate drug discovery. The advances in computational prowess and bioinformatics algorithms have made it possible to handle such huge and complex datasets. In this chapter, we have outlined various computational approaches available for antiviral discovery, with a few case studies.

Keywords Structure-based drug design · QSAR · MD simulation · Structure prediction

17.1 Introduction

Viruses belong to the class of obligate intracellular parasites and can virtually infect all living organisms. A virus is composed of genomic material, which can be either DNA or RNA complexed with an outer protein shell called capsid. Sometimes, a

S. Borkotoky
Department of Biotechnology, Invertis University, Bareilly, Uttar Pradesh, USA

D. Dey (✉)
School of Medicine, University of Maryland Baltimore, Baltimore, MD, USA
e-mail: ddey@som.umaryland.edu

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_17

virus nucleocapsid can also contain an additional lipid membrane containing viral glycoproteins. Manifestation of a viral infection can range from being asymptomatic to potentially life-threatening and even being fatal. Hence, viral diseases have a serious impact on health and economy worldwide causing significant morbidity and mortality (Doms 2016). At the molecular level, infection initiates when an invading virus enters the cells of a susceptible host by attaching itself to certain cellular receptors; for instance, influenza viruses exploit sialic acid on host cells (Sieben et al. 2020), Nipah virus binds ephrin B2 receptor (Negrete et al. 2005), and West Nile virus binds DC-SIGN (Davis et al. 2006), mannose receptor (Cheng et al. 2010), and glycosaminoglycans (Lee et al. 2004). Virus life cycle can be broadly categorized into 5 distinct steps, viz., entry, disassembly and genome release, replication, assembly, and release. During infection, a virus gains control over the cellular replication and protein production machinery to exclusively synthesize multiple copies of the viral genome and proteins. As a result, due to this molecular hijacking, host cells essentially become virus-producing factories, which results in an exponential increase in new infectious particles. New virions then proceed to infect new cells or other hosts and reiterate the process. The severity of infection and extent of metabolic stress often determine the final outcome for the host (Doms 2016). Animals and humans alike are known to be susceptible to a wide range of viruses, for example, avian influenza virus, rift valley fever virus, Nipah virus, new castle disease virus, classical swine fever virus, bovine viral diarrhea virus, foot-and-mouth disease virus, West Nile fever virus, and several others (Brun 2016; Villa et al. 2017). Hence, developing antivirals that either strategically inhibit critical steps in the virus life cycle or provide passive protection is key to abolish or avert a possible infection.

Several small molecule compounds either sourced naturally or chemically synthesized have been shown to possess potent antiviral activity (Table 17.1). These compounds specifically target and inhibit specific steps in the virus life cycle and are therefore an attractive therapeutic option. For instance, natural compounds like flavonoids, terpenoids, triterpenoids, interferons, heparin, chitosan, hyaluronic acid, virocidin, niromycin, feglimycin, and α-defensin 1 are reported to have antiviral activity. Among chemically synthesized molecules, small interfering RNAs (siRNAs), nucleoside analogs, viral protease inhibitors, neuraminidase inhibitors, viral membrane fusion inhibitors, assembly inhibitors, and non-nucleoside reverse transcriptase inhibitors (NNRTIs) are reported. Unlike small molecules, antiviral vaccines utilize the host's own immune system to mount a specific and sustained response against an incoming virus (Villa et al. 2017). Vaccines consist of either attenuated/inactivated virus, specific viral proteins, or their antigenic epitopes, which activate the host immune system to produce neutralizing antibodies that readily mount a strong antiviral response once virus enters the bloodstream. Till date, several antiviral vaccines have been developed against animal viruses and are available commercially (Villa et al. 2017).

Table 17.1 In silico–identified antivirals, their targets, and methodology

Virus	Target	Antiviral	Methodology	Reference
African swine fever virus (ASFV)	AT-rich sequences of virus genome	Congocidine 2 Congocidine 3 Tris-benzimidazole 2	Molecular modeling, molecular docking, all-atom molecular dynamics simulation	Kinyanyi et al. (2019)
	HSP90AB1	Geldanamycin	Protein-protein interaction network analysis, homology modeling, molecular docking, all-atom molecular dynamics simulation	Zhou et al. (2020)
	DNA polymerase	Pentagastrin	Molecular docking, *k*-means clustering, principal component analysis	Choi et al. (2021)
	DNA polymerase	Cangrelor Fostamatinib	*k*-means clustering and principal component analysis–guided unbiased molecular docking	Choi et al. (2021)
Foot and mouth disease virus (FMDV)	3C protease	Adamantane-pyrazole derivatives	Molecular docking	Wassel et al. (2020)
Marburg virus	VP35	CID_11427396	Homology modeling Molecular docking ADME	Quazi et al. (2021)
Ebola virus (EboV)	VP35	ZINC000095486250 NANPDB4048	Structure-based virtual screening Molecular dynamics simulation Naive Bayes classifier Random forest ADME	Darko et al. (2021)
Hendra virus (HeV)	Glycoprotein G	(S)-5-(benzylcarbamoyl)-1-(2-(4-methyl-2-phenylpiperazin-1-yl)-2-oxoethyl)-6-oxo-3,6-dihydropyridin-1-ium-3-ide 5-(cyclohexylcarbamoyl)-1-(2-((2-(3-fluorophenyl)-2-methylpropyl) amino)-2-oxoethyl)-6-oxo-3,6-dihydropyridin-1-ium-3-ide	Structure-guided virtual screening Molecular docking Molecular dynamics simulation	Ahmad et al. (2022)
	RNA-dependent RNA	Nimbolide Kulactone	Structure-based in silico screening	

(continued)

Table 17.1 (continued)

Virus	Target	Antiviral	Methodology	Reference
Japanese encephalitis virus (JEV)	polymerase (RdRp)	Ohchinin Gedunin	Molecular docking molecular dynamics simulation	Dwivedi et al. (2021)
Severe acute respiratory syndrome virus (SARS)-2	Spike	ZINC02111387, ZINC04090608, ZINC02122196 SN00074072	Virtual database screening Molecular docking ADME	Power et al. (2022)
	Spike (RBD)-ACE2 binding interface	Entrectinib	Molecular dynamics simulation Virtual screening	Peralta-Garcia et al. (2021)
	Envelope (E)	Tretinoin Ondansetron	Threading-based molecular modeling Molecular docking, virtual screening MD simulations	Dey et al. (2020)
	RNA-dependent RNA polymerase (RdRp)	Remdesivir Ivermectin	Molecular docking MD simulation	Eweas et al. (2021)

17.2 Challenges of Developing New Antivirals

Development of a new antiviral is a painstakingly long and expensive process with multiple bottlenecks on various fronts. To begin with, research investment in this direction requires sufficient evidence that there is a serious requirement reflected by the number of infections, fatalities, and economic impact caused by the virus infection. This is compounded by the fact that it requires an in-depth understanding of virus biology. Hence, funding is often geared toward economically important viruses, for example, influenza virus. In addition, the preclinical phase witnesses a high rate of attrition and only 30% of marketed compounds break even in terms of the capital invested.

Another important challenge is the emergence of resistant strains, especially when the antiviral is targeted toward an RNA virus, which replicates and mutates at a faster rate than DNA viruses. This is often associated with genetic recombination and reassortment for RNA viruses with a segmented genome like influenza virus, infectious bursal disease virus, and rift valley fever virus. Targeting host proteins or pathways involved would in theory bypass the issue of resistance; however, it might also have serious side-effects for the host.

The fact that some animal viruses are able to infect humans is the direct cause of the third limitation, which is the need for biosafety containment facilities in order to test small molecule antiviral treatments in order to prevent human infections.

Because of the enormous resources and highly skilled employees that are required, these kinds of institutions are only found in select locations around the world. However, virus-like particles, pseudoviruses, or virus mini-replicons can be utilized instead for testing certain small molecule compounds like fusion/entry inhibitors and replication inhibitors.

Antiviral vaccine development, on the contrary, faces challenges like antibody-dependent enhancement (ADE), host genetic variability, and immunodominance, which slows down the overall process significantly. Hence, there is an urgent necessity for streamlining the design and development process for antivirals that would allow for a faster inception-to-clinic transition (Andrei 2021).

17.3 In Silico Methods in Antiviral Drug Research

Antiviral drug discovery involves a multistep process that has two phases of development: preclinical and clinical, respectively. Preclinical phase involves identification, optimization, and validation of candidate small molecules, which are then taken forward for in vitro and in vivo testing. However, there is a lengthy time commitment involved in finding new lead compounds. Computing power, parallel processing, and improved access to super computational facilities have all increased dramatically in recent years, and this has facilitated the creation of computational approaches that significantly quicken the search for effective new antivirals to be tested in vitro and in vivo. In silico intervention has also lowered the rate of failure at later stages. Computational methods for antiviral drug development can be broadly classified as either (1) structure-based or (2) ligand-based (Fig. 17.1).

17.3.1 Structure-Guided Methodologies

This is a very effective and specific approach for identifying new potent antivirals. The primary requirement is a valid three-dimensional (3D) structure of the target

Fig. 17.1 Antiviral drug discovery approaches

viral protein, which is then analyzed computationally for promising small molecule or ligand binding sites that would inactivate the protein. Molecular docking, virtual screening, and molecular dynamics (MD) simulation are commonly used methods for structure-based antiviral drug development. Suitability of a binding partner is determined based on analyses like nature of interactions of the protein and ligand, conformational alterations that occur in target upon interaction with ligand, and binding energy of ligand. Details of the various stages involved are hereby discussed.

17.3.1.1 Structure Determination

First step is a valid structure of the target protein that is obtained experimentally using techniques like X-ray crystallography, nuclear magnetic resonance (NMR), or cryo-electron microscopy (cryoEM). Experimental structures can be obtained freely from databases like the Protein Data Bank (PDB) or European Molecular Database (EMDB), respectively. However, if an experimental structure of the antiviral target is not present, the tertiary structure can also be predicted using protein sequence information. Several online servers or standalone packages are available based on prediction algorithms like homology modeling, threading, ab-initio, or hybrid methods for predicting tertiary structures of proteins. Homology modeling is a very well-known method, which relies on a template protein structure having at least 30% sequence homology to the target protein (Vyas et al. 2012). Some examples include online platforms and/or standalone packages like SWISS-MODEL (Schwede et al. 2003), HHpred (Soding et al. 2005), LOMETS (Wu and Zhang 2007), 3D-JIGSAW (Bates et al. 2001), RaptorX (Kallberg et al. 2014), Phyre2 (Kelley et al. 2015), ROBETTA (Kim et al. 2004), and ESyPred3D (Lambert et al. 2002). The target protein sequence is used as a query in sequence alignment technologies like BLAST to retrieve appropriate templates from databases. The template structure(s) with the highest sequence homology is used for structure determination of the target using either of the three methods, (1) fragment assembly, (2) segment matching, and (3) spatial restraint-based, respectively. Physics-based energy calculations, statistical potentials, or artificial neural networks are utilized for validation of the generated model (Vyas et al. 2012). In case of targets that lack a homologous structure in PDB, threading is a suitable option (Lemer et al. 1995). The inverse protein-folding problem, first proposed by Bowie et al. in 1991, forms the basis of the threading algorithm, which is based on the idea that proteins without sequence homology might have the same fold (Bowie et al. 1991). Threading works by calculating the suitability of each amino acid of the target sequence against a library of template structures using local structural parameters like solvent accessibility, pairwise interactions, and secondary structure. The best-fit templates are utilized for model building, and the lowest energy resultant structure is selected. HHpred (Soding et al. 2005), Phyre, MUSTER, 3D-PSSM (Bates et al. 2001), I-TASSER (Yang et al. 2015), and RaptorX (Kallberg et al. 2014) are some examples of web servers, which utilize protein threading for structure prediction.

Template-free model building platforms like QUARK (Mortuza et al. 2021), 3dProFold (Yousef et al. 2017), and Pep-Fold 2 and 3 (Lamiable et al. 2016) utilize the ab-initio method for structure prediction (Lee et al. 2009). It is a computationally expensive method, which uses an energy function to search all possible conformational space for a given query sequence. Monte Carlo and/or molecular dynamics simulation are utilized as search methods. This results in generation of a number of possible conformations referred to as structure decoys. Energy-based or free energy–based algorithms are used for scoring the decoys to select the final model.

17.3.1.2 Binding Site Determination and Virtual Screening

In order to identify small molecules to inhibit the target protein, it is important to know potential binding sites on the protein surface that can render it inactive. Knowledge of binding sites or cavities can be obtained either experimentally or using binding site prediction servers like CASTp (Binkowski et al. 2003), MetaPocket (Huang 2009), DEPTH (Tan et al. 2013), P2Rank (Krivak and Hoksza 2018), and FunFOLD2 (Roche et al. 2013). Small molecule databases can then be used for identifying potential binders at the specified site using molecular docking, which carries out in silico binding of protein and ligand molecules and scores docked ligands according to their binding affinity. It is one of the most important strategies for antiviral drug discovery. Docking is usually carried out via one of the two methods: (a) simulation method wherein flexibility of the ligand is taken into consideration to identify the optimal docked conformation; and (b) shape complementarity approach wherein complementarity between the protein and ligand is utilized to identify suitable binding pocket in the protein (Forli et al. 2016). The latter approach is computationally much faster. AutoDock Vina (Trott and Olson 2010) and GLIDE (Repasky et al. 2007) are two of the most widely used software packages for molecular docking. Vina uses a united atom scoring function, which considers only heavy atoms, whereas the scoring function of GLIDE is empirical. DrugBank (Wishart et al. 2018), ZINC database (Irwin et al. 2020), ChemDB (Chen et al. 2005), ChEMBL (Gaulton et al. 2012), ChemSpider (Pence and Williams 2010), and IMPPAT (Mohanraj et al. 2018) are some of the databases routinely utilized for in silico screening. Compounds that exhibit high-affinity binding to the target protein are further analyzed for physico-chemical properties, ADMET parameters, and drug-like and pharmacokinetic properties. SwissADME (Daina et al. 2017) is an online platform routinely used for this purpose.

17.3.1.3 Molecular Dynamics Simulation

It is a powerful and versatile computational technique, which can capture the dynamic behavior of biological macromolecules and complexes at the atomic-scale resolution. In a system of interacting molecules, MD simulation calculates movement of atoms using Newton's equation of motion. Potential energies and

forces are calculated using molecular mechanics force fields like AMBER (Case et al. 2005), Gromos96 (Oostenbrink et al. 2004), CHARMM (Vanommeslaeghe and MacKerell Jr. 2015), and OPLS-AA (Robertson et al. 2019). The contribution of molecular dynamics simulation in antiviral research has increased exponentially in recent years. As part of the antiviral drug discovery process, it is frequently used to verify the accuracy of molecular docking predictions (Liu et al. 2018). Simulated trajectories of protein–ligand docked complexes are analyzed for several key parameters like conformational and complex stability, free energies of binding, and nature of interatomic interactions. Because the scoring function from molecular docking software is inaccurate, binding energy calculations using MM-PBSA (molecular mechanics [MM] with Poisson–Boltzmann [PB] and surface area solvation) and MM-GBSA (molecular mechanics [MM] with Generalized Born [GB] and surface area solvation) are particularly important (Genheden and Ryde 2015). Additionally, MD simulations help to understand how mutations in the target protein might alter the binding kinetics of an established antiviral, thus elucidating molecular mechanisms of inhibition and resistance. Popularly used packages for setting up a system for MD simulation are GROMACS (Abraham et al. 2015), Amber (Case et al. 2005), NAMD (Phillips et al. 2020), and Desmond (Bowers et al. 2006).

17.3.2 Ligand-Guided Methodologies

Non-availability of structural information for the target protein makes the process of antiviral drug discovery and development complicated. Thus, all ligand-based antiviral discovery methods are based on the common principle of molecular similarity. The basis of this concept is the observation that molecules that have similar physico-chemical and structural properties exhibit similar activity. Antivirals known to bind the target protein are analyzed for their physico-chemical and structural properties, which provide the blueprint for identifying novel antivirals with similar parameters. In this section, pharmacophore modeling and quantitative structure-activity relationship (QSAR), two widely used methods for ligand-based antiviral discovery, will be discussed.

17.3.2.1 Pharmacophore Modeling

The concept of a pharmacophore was first proposed by Paul Erlrich in 1898 (Guner and Bowen 2014) when he first demonstrated that nerve fibers could bind methylene blue dye. A pharmacophore can be defined as a collection of molecular properties like electrostatics, hydrophobicity, and steric effects, which is critical for binding a specific site on the surface of a target protein molecule in order to modulate its biological functionality. According to IUPAC, a pharmacophore is defined as "an ensemble of steric and electronic features that is essential to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block)

its biological response." Using pharmacophore modeling, one can determine why structurally unrelated compounds are capable of binding the same site on a target protein. Generation of a pharmacophore requires multiple ligands, and conformational search is an important initial step during pharmacophore modeling. Every compound to be screened can have multiple conformations, which can bind the target protein binding site differently. Molecular dynamics simulation and Monte Carlo sampling paired with energy minimization and clustering are preferred techniques for generating all possible conformation for each compound with lowest energy conformer selected for model building (Choudhury and Narahari Sastry 2019; Prathipati et al. 2007).

Selection of ligands for model generation is critical as the process is dependent on structural variations, and ligand type and size. Each molecule is classified into a certain number of features represented as points wherein each feature is a certain interaction with the target site residues. Using the least-square fitting method, feature points of each molecule then are superimposed with others to derive the final model. Commonly used interactions that are considered for model generation are hydrogen bond donors (amines, -OH, acetylenic -CH, -SH, basic amines) and acceptors (atoms with a lone pair of electrons like nitrogen, oxygen, and sulfur), aromatic groups (pi-pi stacking interactions and cation-pi interactions), hydrophobic regions, and positives and negatives (positively and negatively charged ionizable groups), respectively. Additionally, possible steric clash within the binding site is accounted for under exclusion volumes (XVol).

Resultant pharmacophore models are ranked according to their assigned score. High score indicates a higher possibility that compounds under consideration will map to the corresponding model. A structurally diverse dataset is recommended for obtaining a high-quality model (Choudhury and Narahari Sastry 2019). PharmaGist (Schneidman-Duhovny et al. 2008), PharmMapper (Liu et al. 2010), LigandScout (Wolber and Langer 2005), and Phase (Dixon et al. 2006) are some of the popular pharmacophore modeling software packages.

17.3.2.2 Quantitative Structure-Activity Relationship (QSAR)

It is a computational method that helps to establish a quantifiable mathematical correlation between the physio-chemical and structural properties/descriptors of a set of molecules and its experimentally demonstrated biological activity. A QSAR model has the following distinct advantages in terms of small molecule antiviral discovery: (a) it enables prediction of activity for unknown or novel compounds and (b) it helps to shortlist and rank compounds in a database based on their predicted biological activity and (c) optimization of existing compounds for enhanced activity. A QSAR model is a linear equation of the form [Activity = function (property 1, property 2, property N)]. Major steps for successful generation of a QSAR model are as follows: (a) selection of a set of congeneric but chemically diverse compounds, which have shown an experimentally quantifiable biological activity, (b) determination of physio-chemical, electronic, and structural descriptors of

molecules under consideration, (c) determination of property-activity correlation, and (d) model validation.

Success of a QSAR model is highly dependent upon the type of molecular descriptors, which are considered for model building. Descriptors can be of various types and can be classified in two ways. First is based on the nature of the descriptor like hydrophobic (partition coefficient, solubility parameter, distribution coefficient, Hansch's substitution constant, and hydrophobic fragmental constant), steric (Taft's steric parameters, van der Waals radius, van der Waals volume, molar refractivity, and molar volume), electronic (ionization constant, Hammett's constant, Swain and Lupton field parameter, and Taft's inductive (polar) constant), quantum-chemical (net atomic charge, super delocalizability, and energy of the highest and lowest unoccupied molecular orbital), and spatial (radius of gyration, shadow indices, Jur's descriptors, and principal moment of inertia). On the contrary, based on the dimensional nature of the descriptor these can be classified as 1D through 6D, respectively. Due to the large number of available descriptors, statistical methods like multivariable linear regression analysis, principal component analysis, and partial least-square analysis are used to select the right set of descriptors for explaining the corresponding biological activity (Acharya et al. 2011; Evans et al. 2007).

Model validation is the final critical step, which ensures robustness and accuracy of activity predictions for the QSAR model. Validation methods are primarily of two types (a) internal and (b) external. Internal validation methods utilize the compounds that were used for model generation (e.g., least-square fit, cross-validation, chi-squared test, bootstrapping, scrambling, adjusted R2, and root-mean-squared error), whereas external validation considers an external dataset (Acharya et al. 2011; Gramatica 2013). PHASE, CoMFA (comparative molecular field analysis) (Cramer et al. 1988), CoMSIA (comparative similarity indices analysis) (Klebe et al. 1994), and CATALYST (HypoGen module) (Klebe et al. 1994) are programs that are used for 3D QSAR, whereas programs like MOPAC (Stewart 1988), QikProp (https://www.schrodinger.com/qikprop), and ChemAxon (https://www.chemaxon.com) are used for calculating descriptors.

17.3.3 Machine Learning

Artificial intelligence is machine-derived intelligence. Machine learning is a subset of artificial intelligence, which is rapidly becoming a cornerstone in the antiviral drug discovery process. Machine learning algorithms can identify specific patterns in an otherwise non-linear dataset or a dataset with a unique profile. In the field of medicinal chemistry for drug discovery, machine learning has been successfully implemented for prediction and analysis of a compound's biological activity, possible molecular target(s), pharmacokinetics, toxicity, and other physico-chemical parameters. Supervised, unsupervised, and semi-supervised learning are three different types of machine learning algorithms used for antiviral drug discovery. Supervised learning considers training datasets with known labels, which allows

the algorithm to learn how to provide the correct output over time by adjusting for error minimization. On the contrary, unsupervised learning involves pattern recognition in unlabeled datasets. When working on high-dimensional datasets, i.e., datasets having a higher number of features as compared to the number of observations, the algorithm uses the method of dimensional reduction, which facilitates the process of pattern identification and interpretation. Semi-supervised learning combines aspects of both supervised and unsupervised methods. In this technique, only a part of the dataset is labeled (Rebala et al. 2019; Serafim et al. 2021). Some popular machine learning algorithms used in antiviral research are random forest, naive Bayes, support vector machine, and k-means clustering.

Random forest (RF) is a popularly employed supervised machine language algorithm, which uses ensemble learning and is used for performing classification and regression analysis on large datasets having several features. The algorithm consists of a group of uncorrelated decision trees each with a certain prediction. Based on votes assigned, the best fit is determined. Random forest is used for dataset feature selection, classification, and regression in antiviral drug discovery (Breiman 2001; Serafim et al. 2021).

Naive Bayes (NB) algorithm is used for classification of large datasets. It is based on the Bayes theorem that works on conditional probability. Naive Bayes classifiers have been successfully used for classification of large biomedical datasets, which is critical for identification of antiviral targets and excels even in the presence of non-related information or "noise." Another important area of application in drug discovery is predicting potential protein-ligand interactions (Serafim et al. 2021; Webb 2010).

Support vector machine (SVM) is the third popular machine learning algorithm used in antiviral development. It was developed by Vladamir Vapnik and is used as a dataset classifier and regression analysis. The SVM algorithm uses hyperplanes or decision boundaries based on dataset features, which in drug discovery allows the classification between active and inactive compounds. It can also train regression models or rank compounds for activity from different databases for a particular screening parameter (Cortes and Vapnik 1995; Serafim et al. 2021).

k-means clustering algorithm is the most popular unsupervised machine learning algorithm used in antiviral drug discovery. The algorithm typically segregates a given dataset into a certain number of clusters or centroids (k), which is provided as input. The algorithm then selects k distinct data points or clusters. Then, it measures the distance between the first data point in the dataset and the k number of initial clusters and assigns it to the nearest cluster. This is repeated for all the points in the dataset. Next, the means of all clusters are computed, and the data points are measured and clustered using the mean values. This process is repeated a certain number of times. The algorithm then calculates the total variation within each cluster and provides the correct value of k. An elbow plot of variance reduction vs. k also identifies the correct value of k. This algorithm is a valuable tool for antiviral discovery processes like compound selection, virtual screening, QSAR analysis, ADMET prediction (Hartigan and Wong 1979; Serafim et al. 2021).

17.4 Case Studies

17.4.1 Antivirals against African Swine Fever Virus (ASFV)

First identified in Kenya in 1921, African swine fever virus is associated with African swine fever, which infects the swine population like pigs, hogs, bushpigs, and warthogs of many countries worldwide (Galindo and Alonso 2017; Gaudreault et al. 2020). Till date, there are no known approved vaccines or antivirals reported for this viral infection.

In the study, Kinyanyi et al. tested three compounds congocidine 2, congocidine 3, and tris-benzimidazole 2 for their ability to bind conserved AT-rich regions of the ASFV genome using structure-based approach involving a combination of molecular modeling, docking, and all-atom molecular dynamics simulation under explicit solvent conditions. The template structure of an AT-rich decamer was obtained from PDB, and compounds were either obtained from PubChem or prepared from the Internal Coordinate Mechanics algorithm. Docking poses with the best scores for each compound were then subjected to molecular dynamics simulation. Trajectories were analyzed for binding energy values, which revealed all three compounds were high-affinity binders to the AT-rich decamer DNA duplex. In addition, all the compounds exhibited low toxicity with congocidine 2 and 3 exhibiting high water solubility, which is highly preferred for a potential antiviral candidate (Kinyanyi et al. 2019).

In a separate study by Zhu et al. the authors took a different approach and carried out protein-protein interaction network analysis between ASFV proteins and swine proteins to identify potential host targets for developing broad range antivirals. The study identified 21 swine proteins, which were found to interact with ASFV proteins. These proteins were used as input for identifying possible drug candidates from the DrugBank database. A total of 142 drugs were identified. Interestingly, HSP90AA1 (Heat Shock Protein 90 alpha family class A member 1), a central player in ASFV-swine interactions, was found to be targeted by several drugs. In addition, geldanamycin was one of the drugs, which showed specificity toward HSP90AA1 and related HSP90AB1, HSP90B1 swine proteins, all key players during ASFV infections. Homology modeling of HSP90AB1, followed by molecular docking and MD simulation, revealed that geldanamycin could bind with high affinity and form a stable complex with HSP90AB1 (Zhu et al. 2020).

Interestingly, Choi et al. reported the successful use of a hybrid approach, which involved structure-guided molecular docking and the machine learning algorithm k-means clustering along with principal component analysis (PCA) for identification of an FDA-approved drug called pentagastrin against the ASFV DNA polymerase enzyme. This drug was predicted to bind both the free and bound form of the polymerase. Indeed, pentagastrin gave promising results in vitro wherein it inhibited the polymerase activity of the purified enzyme in a dose-dependent manner (Choi et al. 2021b). In a follow-up study, the group identified an underlying inherent limitation of molecular docking software scoring algorithms while carrying out

virtual screening for antiviral discovery. It is based on the fact that the docking score used for ranking is highly influenced by properties of the binding site resulting in bias. The authors utilized k-means clustering and PCA to eliminate the bias, which led to their identification of two new lead antiviral compounds, cangrelor and fostamatinib, which efficiently inhibited ASFV DNA polymerase activity in vitro (Choi et al. 2021a).

17.4.2 Elucidating Molecular Mechanism of 3C Protease Inhibitors of Foot-and-Mouth Disease Virus (FMDV)

Considered as one of the most economically important animal viruses, FMDV affects various members of the Artiodactylous species, sheep, goat, and pigs. Although vaccines for FMDV are commercially available, none provides cross-protection against all the 7 different serotypes and more than 60 subtypes currently are circulating worldwide. Another limitation is the absence of complete protective immunity 7 days' post-vaccine administration (Jamal and Belsham 2013). In a recent study by Wassel et al., the authors tested and validated adamantane-pyrazole derivatives as potent inhibitors of FMDV infection (O serotype) both under in vitro and in vivo conditions. Molecular docking was used as a characterization tool to identify the binding modality of the derivatives against 3C protease of FMDV using the structure-guided approach, thereby providing a glimpse into the molecular mechanism of inhibition. A total of 9 derivatives were tested out of which only 3 (compounds 6a, 6b, and 6c) exhibited potent antiviral activity. Molecular docking analysis predicted compound 6a as the strongest binder (−20 kcal/mol). Binding was established between Lys 5, Arg 68 of FMDV 3C protease and carbonyl, anisidine, pyrazole moiety, and phenyl group of 6a (Wassel et al. 2020).

17.4.3 Novel VP35 Inhibitors Against Marburg (MARV) and Ebola (EBOV)

Viral hemorrhagic fevers are caused by members of the *filoviridae* family. Marburg and Ebola are well-known members of this family, which, besides humans, infect non-human primates like chimpanzees, gorillas, rhesus monkeys, cynomolgus monkeys, baboons, newborn mice, African green monkeys, baboons, and vervet monkeys (Takada 2012). Two reports by Qazi et al. and Darko et al. have reported potential antiviral candidates targeted against the VP35 protein of Marburg and Ebola viruses, respectively.

Using a combination of molecular modeling, virtual database screening, docking, and ADME analysis, Quazi et al. identified three PubChem database candidates, CID_5477931, CID_ 3,007,938, and CID_11427396, as potential inhibitors of

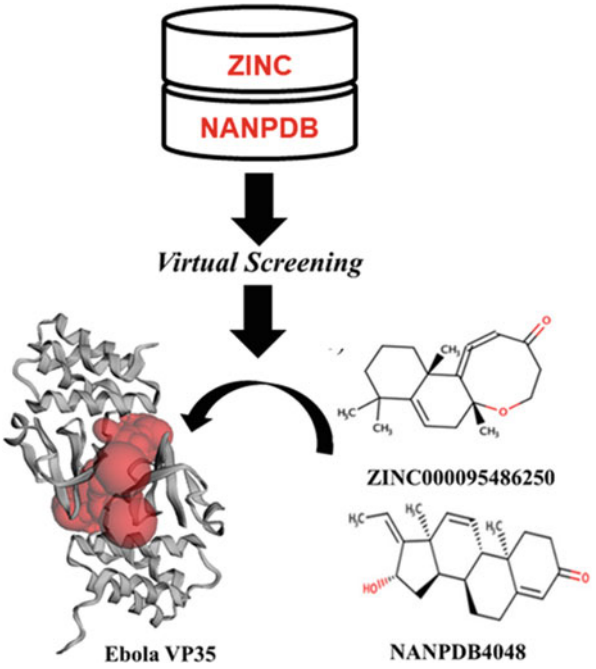

Fig. 17.2 Schematic representation of in silico antiviral discovery against Ebola virus VP35

VP35. Database screening was conducted in two stages. In stage 1, the drug FGI-103 was used as input to screen for structurally similar compounds. In stage 2, output of stage 1 was subjected to Pfizer cross-validation and Lipinski's rule of five analysis to identify drug-like molecules. This approach gave three final hits, which were then subjected to molecular docking with a predicted structure of VP35 using homology modeling. Binding site selected for docking was the P1 site, shown previously to bind FGI-103. Out of the 3 selected candidates, CID_11427396 showed the lowest binding energy with VP35 and formed a stable complex involving hydrogen bonds and van der Waals interactions. All candidates further also passed ADME analysis (Quazi et al. 2021). On the contrary, Darko et al. carried out an extensive in silico analysis (Fig. 17.2) for identifying novel antivirals, which could specifically bind the interferon inhibitory domain of EBOV VP35. In this study, the authors adopted a hybrid strategy of structure-based virtual screening, molecular dynamics simulation, and naive Bayes–based machine learning classifier algorithm to identify four potential antivirals NANPDB4048, NANPDB2412, ZINC000095486250, and NANPDB2476 against VP35.

The template structure of the VP35 domain was obtained from PDB. A total of 7675 compounds were screened using a ADMET predictor. The binding site was identified and characterized using the CASTp web server, which was used for molecular docking of 1470 screened compounds using AutoDock Vina. 94 ligands were selected having binding energy values greater than −8.1 kcal/mol cutoff. These ligands were further analyzed and screened using ADME profiling, protein-ligand

interactions, and various ligand efficiency metrics. MD simulation followed by MM-PBSA analysis of the top 4 compounds in complex with VP35 revealed that ZINC000095486250 had the lowest binding energy wherein van der Waals energy was the dominant contributor toward the total binding energy. Naive Bayes classifier and random forest algorithm were utilized for predicting biological activity and IC50 values, respectively. All selected compounds were inhibitors of DNA polymerase I, Ebola virus proteins, transcription factors, and RNA synthesis with NANPDB4048 having the best predicted IC50 value of 3.35 uM (Darko et al. 2021).

17.4.4 Hendra Virus (HeV) G Glycoprotein Targeting Antivirals

First discovered in 1994, HeV is a highly pathogenic member of the *Paramyxoviridae* family. HeV infection has a high mortality rate and is known to naturally infect horses, pigs, fruit bats, and humans (Field et al. 2010; Wang et al. 2021). No antiviral or vaccine has been approved till date. Ahmad et al. carried out structure-guided virtual screening against the G glycoprotein of HeV to identify potential binders. Molecular docking and MD simulation were further carried out to identify and validate the strongest binders to both free and bound forms of G glycoprotein. The G glycoprotein structure in complex with the host ephrin B receptor was obtained from PDB. The Asinex Biodesign database was used for virtual screening. A total of 3560 candidates were subjected to docking with free G glycoprotein structure using AutoDock Vina (Trott and Olson 2010) post-screening for various pharmacokinetic and drug-likeness parameters. This was followed by docking the best compounds from the previous docking step with the ephrin B bound structure. Two compounds, namely, (S)-5-(benzylcarbamoyl)-1-(2-(4-methyl-2-phenylpiperazin-1-yl)-2-oxoethyl)-6-oxo-3,6-dihydropyridin-1-ium-3-ide and 5-(cyclohexylcarbamoyl)-1-(2-((2-(3-fluorophenyl)-2-methylpropyl) amino)-2-oxoethyl)-6-oxo-3,6-dihydropyridin-1-ium-3-ide, were selected as potential candidates. MD simulation was carried out using AMBER (Case et al. 2005). Binding energy analysis was done by MM-PBSA/MM-GBSA (Genheden and Ryde 2015). The analysis showed that both free and bound forms of the complexes were stable. Majority of binding energy contribution came from van der Waals interaction energy. Till date, this is the first report of potential antivirals, which can be tested against HeVs (Ahmad et al. 2022).

17.4.5 Inhibitors of Japanese Encephalitis Virus (JEV) RNA-Dependent RNA Polymerase (RdRp)

Japanese encephalitis virus is a re-emerging member of the flavivirus family known to cause potentially fatal neurological complications due to its ability to infect brain

tissues leading to inflammation. Animals like donkeys, pigs, and horses are susceptible, whereas dogs, cats, goats, sheep, and cattle can have asymptomatic infection. Dwivedi et al. carried out a comprehensive study to identify natural bioflavonoids reported in *Azadirachta indica*, which could inhibit JEV RdRp, a key molecular player in the virus life cycle. The strategy was based on structure-based in silico screening, molecular docking, and molecular dynamics simulation. The template structure of JEV RdRp was obtained from PDB, and 43 neem bioflavonoids were obtained from PubChem. The screening was done using the MTi-OpenScreen server (Labbe et al. 2015) at the RdRp binding site reported in literature. Top 4 compounds, i.e., Nimbolide, Kulactone, Ohchinin acetate, and Gedunin, were selected and redocked using AutoDock Vina followed by all-atom molecular dynamics simulations. Trajectory and binding energy analyses revealed that all compounds formed stable complexes with JEV RdRp and van der Waals interactions were the major interactive force between the ligands and the polymerase (Dwivedi et al. 2021).

17.5 In Silico Antiviral Discovery During the COVID-19 Pandemic

The recent global SARS-2 pandemic has reinstated the important role in silico techniques play in efficiently and effectively speeding up the process of antiviral discovery. Ever since its start, the biomedical literature has been flooded with articles that propose in silico–identified compounds as potential antivirals against the SARS-2 coronavirus (Khan et al. 2021a, b; Pant et al. 2021; Islam et al. 2021; Kundu et al. 2021; Power et al. 2022; Dey et al. 2020; Bayat et al. 2021; Eweas et al. 2021; Peralta-Garcia et al. 2021). Several of these compounds have been validated in vitro and in vivo (Fig. 17.3). This section focuses on some of these identified compounds and in silico techniques utilized.

The viral spike protein is a key target for inhibitors due to its primary role in virus entry (Jackson et al. 2022). Power et al. reported the identification of 4 key natural

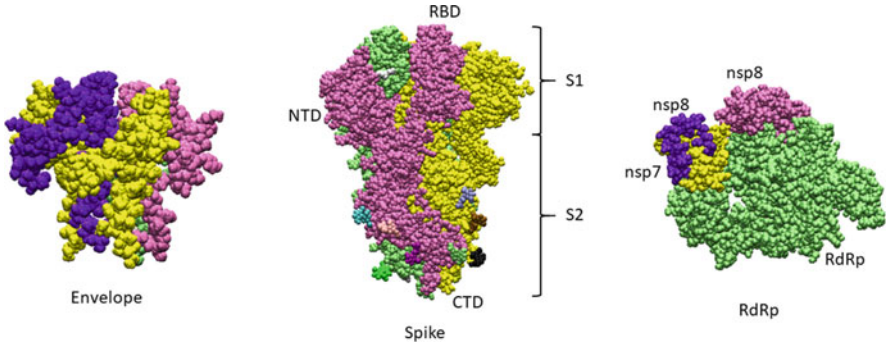

Fig. 17.3 3D structures of the SARS-2 antiviral targets

compounds (ZINC02111387, ZINC04090608, ZINC02122196, and SN00074072) using a top-down approach of virtual database screening, molecular docking (AutoDock Vina), and ADME analysis followed by in vitro cell-based validation. A total of 52,72,09 compounds were screened (Power et al. 2022). Repurposing of FDA-approved drugs has also given promising leads. The host ACE-2 receptor is the binding partner of SARS-2 spike (Jackson et al. 2022). Peralta-Garcia et al. utilized long-scale molecular dynamics simulation to identify critical residues at the binding interface of the spike-receptor binding domain (RBD) and ACE-2 receptor wherein residue-pairs with >90% contact frequency were considered. A virtual screen of 5849 FDA-approved compounds using the simulation-derived contact map was executed. Out of the top 5 hits considered, entrectinib was selected as the top candidate based on its observed inhibitory effects in human lung tissue cells (Peralta-Garcia et al. 2021). Tretinoin and ondansetron are two such drugs identified using structure-guided approach, which involved in silico structure determination, molecular docking, virtual screening, and MD simulations (Dey et al. 2020). Both drugs were predicted to be effective inhibitors of the SARS-2 E ion channel protein. A study involving more than 10,000 patients showed a marked reduction in morbidity and mortality when ondansetron was administered within 48 h post-infection (Bayat et al. 2021). Tretinoin on the contrary has also shown promise, and 2 clinical trials are currently underway (National Library of Medicine [NLM] (2020), NCT04353180, and National Library of Medicine [NLM] (2021), NCT05002530). Eweas et al., reported the identification of remdesivir and ivermectin as potential antivirals against the SARS-2 RNA-dependent RNA polymerase (RdRp) using a combination of molecular docking and MD simulation approach (Eweas et al. 2021). This was eventually confirmed by in vitro and in vivo studies (Biber et al. 2022; Wang et al. 2020; Williamson et al. 2020). Currently, intravenous formulation of remdesivir is approved by the FDA for treatment of COVID-19 infection in adults and pediatric patients (https://www.covid19treatmentguidelines.nih.gov/therapies/antiviral-therapy/remdesivir/). However, a growing concern among epidemiologists and the biomedical research community is the reverse spillover of SARS-2 variants from humans into animals, thereby creating room for emergence of novel mutations, which could quickly render existing therapeutics ineffective. Although preliminary studies have shown no instance of SARS-2 infection in some local bat populations (Dakroub et al. 2022), a potential animal reservoir has been confirmed in the white-tailed deer population of North America (Hale et al. 2022). As of now, three lineages B.1.2, B.1.582, and B.1.596 have been identified. Hence, a future spillover of novel and perhaps deadlier animal SARS-2 variants into the human population is a strong possibility. Hence, further robust and accurate prediction tools need to be developed. A promising advancement is the introduction of a tensor decomposition (TD)–based unsupervised feature extraction (FE) method (Taguchi 2017). It is an unsupervised method, which can be applied to discover antiviral drugs when there is no prior knowledge of an effective antiviral. It has been applied successfully for identification of antivirals against using a dataset of differentially expressed gene profiles in several lung carcinoma cell lines infected with SARS-2 (Taguchi and Turki 2020).

17.6 Concluding Remarks

Antiviral drug discovery and development is one of the most challenging areas of present-day biomedical research due to stringent requirements of safety, efficacy, and high rates of failure, which extends the timeline for commercialization of any new antiviral significantly. While the stringency is an absolute necessity, expediting the preclinical stage, particularly lead compound identification and optimization is a desirable cost-effective option. Bioinformatics tools and methods in recent years have evolved dramatically, which makes working with large and complicated datasets a reality. In addition, development of molecular modeling techniques and artificial intelligence allow antiviral drug discovery to increase efficiency by several folds at a fraction of the time. Introduction of the revolutionary AI-based protein structure prediction tool ALPHA-FOLD (Jumper et al. 2021) is particularly encouraging since it now allows researchers to expedite structure-based antiviral discovery for viruses whose protein structures have not yet been solved by conventional wet laboratory techniques like crystallography, NMR, or cryoEM. This results in prospective antiviral candidates being more likely to be successful downstream of the antiviral drug development pipeline. It is envisaged that the process of discovering antivirals would become even more reliable and robust in the years to come as a result of subsequent technological breakthroughs.

References

Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. Software X 1-2:19–25. https://doi.org/10.1016/j.softx.2015.06.001

Acharya C, Coop A, Polli JE, Mackerell AD Jr (2011) Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. Curr Comput Aided Drug Des 7(1):10–22. https://doi.org/10.2174/157340911793743547

Ahmad F, Albutti A, Tariq MH, Din G, Qamar MTU, Ahmad S (2022) Discovery of potential antiviral compounds against Hendra virus by targeting its receptor-binding protein (G) using computational approaches. Molecules 27(2):554. https://doi.org/10.3390/molecules27020554

Andrei G (2021) Vaccines and antivirals: grand challenges and great opportunities. Front Virol 1: 666548. https://doi.org/10.3389/fviro.2021.666548

Bates PA, Kelley LA, MacCallum RM, Sternberg MJ (2001) Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. Proteins 5: 39–46. https://doi.org/10.1002/prot.1168

Bayat V, Ryono R, Phelps S, Geis E, Sedghi F, Etminani P, Holodniy M (2021) Reduced mortality with ondansetron use in SARS-CoV-2-infected inpatients. Open Forum Infect Dis 8(7):ofab336. https://doi.org/10.1093/ofid/ofab336

Biber A, Harmelin G, Lev D, Ram L, Shaham A, Nemet I, Kliker L, Erster O, Mandelboim M, Schwartz E (2022) The effect of ivermectin on the viral load and culture viability in early treatment of non-hospitalized patients with mild COVID-19 - A double-blind, randomized placebo-controlled trial. Int J Infect Dis 122:733–740. https://doi.org/10.1016/j.ijid.2022.07.003

Binkowski TA, Naghibzadeh S, Liang J (2003) CASTp: computed atlas of surface topography of proteins. Nucleic Acids Res 31(13):3352–3355. https://doi.org/10.1093/nar/gkg512

Bowers KJ, Chow DE, Xu H, Dror RO, Eastwood MP, Gregersen BA, Shaw DE (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters. Paper presented at the SC '06: proceedings of the 2006 ACM/IEEE conference on supercomputing

Bowie JU, Luthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. Science 253(5016):164–170. https://doi.org/10.1126/science.1853201

Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324

Brun A (2016) Vaccines and vaccination for veterinary viral diseases: A general overview. Methods Mol Biol 1349:1–24. https://doi.org/10.1007/978-1-4939-3008-1_1

Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, Woods RJ (2005) The Amber biomolecular simulation programs. J Comput Chem 26(16):1668–1688. https://doi.org/10.1002/jcc.20290

Chen J, Swamidass SJ, Dou Y, Bruand J, Baldi P (2005) ChemDB: a public database of small molecules and related chemoinformatics resources. Bioinformatics 21(22):4133–4139. https://doi.org/10.1093/bioinformatics/bti683

Cheng G, Cox J, Wang P, Krishnan MN, Dai J, Qian F, Fikrig E (2010) A C-type lectin collaborates with a CD45 phosphatase homolog to facilitate West Nile virus infection of mosquitoes. Cell 142(5):714–725. https://doi.org/10.1016/j.cell.2010.07.038

Choi J, Tark D, Lim YS, Hwang SB (2021a) Identification of African swine fever virus inhibitors through high performance virtual screening using machine learning. Int J Mol Sci 22(24):13414. https://doi.org/10.3390/ijms222413414

Choi J, Yun JS, Song H, Shin YK, Kang YH, Munashingha PR, Hwang SB (2021b) Prediction of African swine fever virus inhibitors by molecular docking-driven machine learning models. Molecules 26(12):3592. https://doi.org/10.3390/molecules26123592

Choudhury C, Narahari Sastry G (2019) Pharmacophore modelling and screening: concepts, recent developments and applications in rational drug design. In: Mohan CG (ed) Structural bioinformatics: applications in preclinical drug discovery process. Springer International Publishing, Cham, pp 25–53

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297. https://doi.org/10.1007/BF00994018

Cramer RD, Patterson DE, Bunce JD (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 110(18):5959–5967. https://doi.org/10.1021/ja00226a005

Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep 7:42717. https://doi.org/10.1038/srep42717

Dakroub H, Russo D, Cistrone L, Serra F, Fusco G, De Carlo E, Amoroso MG (2022) A first assessment of SARS-CoV-2 circulation in bats of central-southern Italy. Pathogens (Basel, Switzerland) 11(7):742. https://doi.org/10.3390/pathogens11070742

Darko LKS, Broni E, Amuzu DSY, Wilson MD, Parry CS, Kwofie SK (2021) Computational study on potential novel anti-Ebola virus protein VP35 natural compounds. Biomedicine 9(12):1796. https://doi.org/10.3390/biomedicines9121796

Davis CW, Nguyen HY, Hanna SL, Sanchez MD, Doms RW, Pierson TC (2006) West Nile virus discriminates between DC-SIGN and DC-SIGNR for cellular attachment and infection. J Virol 80(3):1290–1301. https://doi.org/10.1128/JVI.80.3.1290-1301.2006

Dey D, Borkotoky S, Banerjee M (2020) *In silico* identification of Tretinoin as a SARS-CoV-2 envelope (E) protein ion channel inhibitor. Comput Biol Med 127:104063. https://doi.org/10.1016/j.compbiomed.2020.104063

Dixon SL, Smondyrev AM, Rao SN (2006) PHASE: a novel approach to pharmacophore modeling and 3D database searching. Chem Biol Drug Des 67(5):370–372. https://doi.org/10.1111/j.1747-0285.2006.00384.x

Doms RW (2016) Chapter 3 - basic concepts: a step-by-step guide to viral infection. In: Katze MG, Korth MJ, Law GL, Nathanson N (eds) Viral pathogenesis, 3rd edn. Academic Press, Boston, pp 29–40

Dwivedi VD, Singh A, El-Kafraway SA, Alandijany TA, Faizo AA, Bajrai LH, Azhar EI (2021) Mechanistic insights into the Japanese encephalitis virus RNA dependent RNA polymerase protein inhibition by bioflavonoids from Azadirachta indica. Sci Rep 11(1):18125. https://doi.org/10.1038/s41598-021-96917-0

Evans DA, Doman TN, Thorner DA, Bodkin MJ (2007) 3D QSAR methods: phase and Catalyst compared. J Chem Inf Model 47(3):1248–1257. https://doi.org/10.1021/ci7000082

Eweas AF, Alhossary AA, Abdel-Moneim AS (2021) Molecular docking reveals Ivermectin and Remdesivir as potential repurposed drugs against SARS-CoV-2. Front Microbiol 11:592908. https://doi.org/10.3389/fmicb.2020.592908

Field H, Schaaf K, Kung N, Simon C, Waltisbuhl D, Hobert H, Lovell D (2010) Hendra virus outbreak with novel clinical features, Australia. Emerg Infect Dis 16(2):338–340. https://doi.org/10.3201/eid1602.090780

Forli S, Huey R, Pique ME, Sanner MF, Goodsell DS, Olson AJ (2016) Computational protein-ligand docking and virtual drug screening with the AutoDock suite. Nat Protoc 11(5):905–919. https://doi.org/10.1038/nprot.2016.051

Galindo I, Alonso C (2017) African swine fever virus: a review. Viruses 9(5):103. https://doi.org/10.3390/v9050103

Gaudreault NN, Madden DW, Wilson WC, Trujillo JD, Richt JA (2020) African swine fever virus: an emerging DNA arbovirus. Front Vet Sci 7:215. https://doi.org/10.3389/fvets.2020.00215

Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Overington JP (2012) ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 40(Database issue):D1100–D1107. https://doi.org/10.1093/nar/gkr777

Genheden S, Ryde U (2015) The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov 10(5):449–461. https://doi.org/10.1517/17460441.2015.1032936

Gramatica P (2013) On the development and validation of QSAR models. Methods Mol Biol 930:499–526. https://doi.org/10.1007/978-1-62703-059-5_21

Guner OF, Bowen JP (2014) Setting the record straight: the origin of the pharmacophore concept. J Chem Inf Model 54(5):1269–1283. https://doi.org/10.1021/ci5000533

Hale VL, Dennis PM, McBride DS, Nolting JM, Madden C, Huey D, Ehrlich M, Grieser J, Winston J, Lombardi D, Gibson S, Saif L, Killian ML, Lantz K, Tell RM, Torchetti M, Robbe-Austerman S, Nelson MI, Faith SA, Bowman AS (2022) SARS-CoV-2 infection in free-ranging white-tailed deer. Nature 602(7897):481–486. https://doi.org/10.1038/s41586-021-04353-x

Hartigan JA, Wong MA (1979) Algorithm AS 136: A K-means clustering algorithm. J R Stat Soc Ser C (Appl Stat) 28(1):100–108. https://doi.org/10.2307/2346830

Huang B (2009) MetaPocket: a meta approach to improve protein ligand binding site prediction. OMICS 13(4):325–330. https://doi.org/10.1089/omi.2009.0045

Irwin JJ, Tang KG, Young J, Dandarchuluun C, Wong BR, Khurelbaatar M, Sayle RA et al (2020) ZINC20-A free Ultralarge-scale chemical database for ligand discovery. J Chem Inf Model 60(12):6065–6073. https://doi.org/10.1021/acs.jcim.0c00675

Islam R, Parves MR, Paul AS, Uddin N, Rahman MS, Mamun AA, Hossain MN, Ali MA, Halim MA (2021) A molecular modeling approach to identify effective antiviral phytochemicals against the main protease of SARS-CoV-2. J Biomol Struct Dyn 39(9):3213–3224. https://doi.org/10.1080/07391102.2020.1761883

Jackson CB, Farzan M, Chen B, Choe H (2022) Mechanisms of SARS-CoV-2 entry into cells. Nat Rev Mol Cell Biol 23(1):3–20. https://doi.org/10.1038/s41580-021-00418-x

Jamal SM, Belsham GJ (2013) Foot-and-mouth disease: past, present and future. Vet Res 44:116. https://doi.org/10.1186/1297-9716-44-116

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Hassabis D (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596(7873):583–589. https://doi.org/10.1038/s41586-021-03819-2

Kallberg M, Margaryan G, Wang S, Ma J, Xu J (2014) RaptorX server: a resource for template-based protein structure modeling. Methods Mol Biol 1137:17–27. https://doi.org/10.1007/978-1-4939-0366-5_2

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10(6):845–858. https://doi.org/10.1038/nprot.2015.053

Khan SA, Zia K, Ashraf S, Uddin R, Ul-Haq Z (2021a) Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach. J Biomol Struct Dyn 39(7):2607–2616. https://doi.org/10.1080/07391102.2020.1751298

Khan RJ, Jha RK, Amera GM, Jain M, Singh E, Pathak A, Singh RP, Muthukumaran J, Singh AK (2021b) Targeting SARS-CoV-2: a systematic drug repurposing approach to identify promising inhibitors against 3C-like proteinase and 2'-O-ribose methyltransferase. J Biomol Struct Dyn 39(8):2679–2692. https://doi.org/10.1080/07391102.2020.1753577

Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 32(Web Server issue), W526-531:W526. https://doi.org/10.1093/nar/gkh468

Kinyanyi D, Amwayi P, Wamalwa M, Obiero G (2019) Comparative *in silico* study of congocidine congeners as potential inhibitors of African swine fever virus. PLoS One 14(8):e0221175. https://doi.org/10.1371/journal.pone.0221175

Klebe G, Abraham U, Mietzner T (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. J Med Chem 37(24):4130–4146. https://doi.org/10.1021/jm00050a010

Krivak R, Hoksza D (2018) P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. J Cheminform 10(1):39. https://doi.org/10.1186/s13321-018-0285-8

Kundu D, Selvaraj C, Singh SK, Dubey VK (2021) Identification of new anti-nCoV drug chemical compounds from Indian spices exploiting SARS-CoV-2 main protease as target. J Biomol Struct Dyn 39(9):3428–3434. https://doi.org/10.1080/07391102.2020.1763202

Labbe CM, Rey J, Lagorce D, Vavrusa M, Becot J, Sperandio O, Miteva MA (2015) MTiOpenScreen: a web server for structure-based virtual screening. Nucleic Acids Res 43(W1):W448–W454. https://doi.org/10.1093/nar/gkv306

Lambert C, Leonard N, De Bolle X, Depiereux E (2002) ESyPred3D: prediction of proteins 3D structures. Bioinformatics 18(9):1250–1256. https://doi.org/10.1093/bioinformatics/18.9.1250

Lamiable A, Thevenet P, Rey J, Vavrusa M, Derreumaux P, Tuffery P (2016) PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. Nucleic Acids Res 44(W1):W449–W454. https://doi.org/10.1093/nar/gkw329

Lee E, Hall RA, Lobigs M (2004) Common E protein determinants for attenuation of glycosaminoglycan-binding variants of Japanese encephalitis and West Nile viruses. J Virol 78(15):8271–8280. https://doi.org/10.1128/JVI.78.15.8271-8280.2004

Lee J, Wu S, Zhang Y (2009) Ab initio protein structure prediction. In: Rigden DJ (ed) From protein structure to function with bioinformatics. Springer Netherlands, Dordrecht, pp 3–25

Lemer CM, Rooman MJ, Wodak SJ (1995) Protein structure prediction by threading methods: evaluation of current techniques. Proteins 23(3):337–355. https://doi.org/10.1002/prot.340230308

Liu X, Ouyang S, Yu B, Liu Y, Huang K, Gong J, Jiang H (2010) PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. Nucleic Acids Res 38(Web Server issue):W609–W614. https://doi.org/10.1093/nar/gkq300

Liu X, Shi D, Zhou S, Liu H, Liu H, Yao X (2018) Molecular dynamics simulations and novel drug discovery. Expert Opin Drug Discov 13(1):23–37. https://doi.org/10.1080/17460441.2018.1403419

Mohanraj K, Karthikeyan BS, Vivek-Ananth RP, Chand RPB, Aparna SR, Mangalapandi P, Samal A (2018) IMPPAT: A curated database of Indian medicinal plants, phytochemistry and therapeutics. Sci Rep 8(1):4329. https://doi.org/10.1038/s41598-018-22631-z

Mortuza SM, Zheng W, Zhang C, Li Y, Pearce R, Zhang Y (2021) Improving fragment-based ab initio protein structure assembly using low-accuracy contact-map predictions. Nat Commun 12(1):5011. https://doi.org/10.1038/s41467-021-25316-w

National Library of Medicine (U.S.) (2020) Assessment the Activity Value of Isotretinoin (13-Cis-Retinoic Acid) in the Treatment of COVID-19 (Isotretinoin in Treatment of COVID-19) (Randomized) (Isotretinoin). Identifier: NCT04353180. https://clinicaltrials.gov/ct2/show/NCT04353180

National Library of Medicine (U.S.) (2021) Investigating the Potential Role of Aerosolized Retinoic Acid, a Potent Vitamin A Metabolite for Treating COVID-19 Anosmia and Retinoic Acid Insufficiency. A Novel Approach for Regaining Sense of Smell. Identifier: NCT05002530. https://clinicaltrials.gov/ct2/show/NCT04353180

Negrete OA, Levroney EL, Aguilar HC, Bertolotti-Ciarlet A, Nazarian R, Tajyar S, Lee B (2005) EphrinB2 is the entry receptor for Nipah virus, an emergent deadly paramyxovirus. Nature 436(7049):401–405. https://doi.org/10.1038/nature03838

Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem 25(13):1656–1676. https://doi.org/10.1002/jcc.20090

Pant S, Singh M, Ravichandiran V, Murty U, Srivastava HK (2021) Peptide-like and small-molecule inhibitors against Covid-19. J Biomol Struct Dyn 39(8):2904–2913. https://doi.org/10.1080/07391102.2020.1757510

Pence HE, Williams A (2010) ChemSpider: an online chemical information resource. J Chem Educ 87(11):1123–1124. https://doi.org/10.1021/ed100697w

Peralta-Garcia A, Torrens-Fontanals M, Stepniewski TM, Grau-Expósito J, Perea D, Ayinampudi V, Waldhoer M, Zimmermann M, Buzón MJ, Genescà M, Selent J (2021) Entrectinib-A SARS-CoV-2 inhibitor in human lung tissue (HLT) cells. Int J Mol Sci 22(24):13592. https://doi.org/10.3390/ijms222413592

Phillips JC, Hardy DJ, Maia JDC, Stone JE, Ribeiro JV, Bernardi RC, Tajkhorshid E (2020) Scalable molecular dynamics on CPU and GPU architectures with NAMD. J Chem Phys 153(4):044130. https://doi.org/10.1063/5.0014475

Power H, Wu J, Turville S, Aggarwal A, Valtchev P, Schindeler A, Dehghani F (2022) Virtual screening and in vitro validation of natural compound inhibitors against SARS-CoV-2 spike protein. Bioorg Chem 119:105574. https://doi.org/10.1016/j.bioorg.2021.105574

Prathipati P, Dixit A, Saxena KA (2007) Computer-aided drug design: integration of structure-based and ligand-based approaches in drug design. Curr Comput Aided Drug Des 3(2):133–148. https://doi.org/10.2174/157340907780809516

Quazi S, Malik J, Capuzzo AM, Suman KS, Haider Z (2021) In-silico structural and molecular docking-based drug discovery against viral protein (VP35) of Marburg virus: a potent agent of MAVD. bioRxiv:2021–2002. https://doi.org/10.1101/2021.02.09.430405

Rebala G, Ravi A, Churiwala S (2019) Machine learning definition and basics. In: An introduction to machine learning. Springer International Publishing, Cham, pp 1–17

Repasky MP, Shelley M, Friesner RA (2007) Flexible ligand docking with Glide. Curr Protoc Bioinformatics Chapter 8, Unit 8:12. https://doi.org/10.1002/0471250953.bi0812s18

Robertson MJ, Qian Y, Robinson MC, Tirado-Rives J, Jorgensen WL (2019) Development and testing of the OPLS-AA/M force Field for RNA. J Chem Theory Comput 15(4):2734–2742. https://doi.org/10.1021/acs.jctc.9b00054

Roche DB, Buenavista MT, McGuffin LJ (2013) The FunFOLD2 server for the prediction of protein-ligand interactions. Nucleic Acids Res 41(Web Server issue):W303–W307. https://doi.org/10.1093/nar/gkt498

Schneidman-Duhovny D, Dror O, Inbar Y, Nussinov R, Wolfson HJ (2008) PharmaGist: a webserver for ligand-based pharmacophore detection. Nucleic Acids Res 36(Web Server issue):W223–W228. https://doi.org/10.1093/nar/gkn187

Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 31(13):3381–3385. https://doi.org/10.1093/nar/gkg520

Serafim MSM, Dos Santos Junior VS, Gertrudes JC, Maltarollo VG, Honorio KM (2021) Machine learning techniques applied to the drug design and discovery of new antivirals: a brief look over the past decade. Expert Opin Drug Discov 16(9):961–975. https://doi.org/10.1080/17460441.2021.1918098

Sieben C, Sezgin E, Eggeling C, Manley S (2020) Influenza A viruses use multivalent sialic acid clusters for cell binding and receptor activation. PLoS Pathog 16(7):e1008656. https://doi.org/10.1371/journal.ppat.1008656

Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33(Web Server issue):W244–W248. https://doi.org/10.1093/nar/gki408

Stewart, J. J. (1988). MOPAC Manual: A General Molecular Orbital Package. Retrieved from http://openmopac.net/

Taguchi YH (2017) Tensor decomposition-based unsupervised feature extraction applied to matrix products for multi-view data processing. PLoS One 12(8):e0183933. https://doi.org/10.1371/journal.pone.0183933

Taguchi YH, Turki T (2020) A new advanced *in silico* drug discovery method for novel coronavirus (SARS-CoV-2) with tensor decomposition-based unsupervised feature extraction. PLoS One 15(9):e0238907. https://doi.org/10.1371/journal.pone.0238907

Takada A (2012) Filovirus tropism: cellular molecules for viral entry. Front Microbiol 3:34. https://doi.org/10.3389/fmicb.2012.00034

Tan KP, Nguyen TB, Patel S, Varadarajan R, Madhusudhan MS (2013) Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. Nucleic Acids Res 41(Web Server issue):W314–W321. https://doi.org/10.1093/nar/gkt503

Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31(2):455–461. https://doi.org/10.1002/jcc.21334

Vanommeslaeghe K, MacKerell AD Jr (2015) CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. Biochim Biophys Acta 1850(5):861–871. https://doi.org/10.1016/j.bbagen.2014.08.004

Villa TG, Feijoo-Siota L, Rama JLR, Ageitos JM (2017) Antivirals against animal viruses. Biochem Pharmacol 133:97–116. https://doi.org/10.1016/j.bcp.2016.09.029

Vyas VK, Ukawala RD, Ghate M, Chintha C (2012) Homology modeling a fast tool for drug discovery: current perspectives. Indian J Pharm Sci 74(1):1–17. https://doi.org/10.4103/0250-474X.102537

Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, Shi Z, Hu Z, Zhong W, Xiao G (2020) Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Res 30(3):269–271. https://doi.org/10.1038/s41422-020-0282-0

Wang J, Anderson DE, Halpin K, Hong X, Chen H, Walker S, Williams DT (2021) A new Hendra virus genotype found in Australian flying foxes. Virol J 18(1):197. https://doi.org/10.1186/s12985-021-01652-7

Wassel MMS, Gamal Eldin WM, Ragab A, Elhag Ali GAM, Ammar Y YA (2020) Antiviral activity of Adamantane-Pyrazole derivatives against foot and mouth disease virus infection

in vivo and in vitro with molecular docking study. J Appl Vet Sci 5(4):37–46. https://doi.org/10.21608/javs.2020.118001

Webb GI (2010) Naïve Bayes. In: Sammut C, Webb GI (eds) Encyclopedia of machine learning. Springer US, Boston, MA, pp 713–714

Williamson BN, Feldmann F, Schwarz B, Meade-White K, Porter DP, Schulz J, van Doremalen N, Leighton I, Yinda CK, Pérez-Pérez L, Okumura A, Lovaglio J, Hanley PW, Saturday G, Bosio CM, Anzick S, Barbian K, Cihlar T, Martens C, Scott DP, de Wit E et al (2020) Clinical benefit of remdesivir in rhesus macaques infected with SARS-CoV-2. Nature 585(7824):273–276. https://doi.org/10.1038/s41586-020-2423-5

Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Wilson M (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 46(D1):D1074–D1082. https://doi.org/10.1093/nar/gkx1037

Wolber G, Langer T (2005) LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J Chem Inf Model 45(1):160–169. https://doi.org/10.1021/ci049885e

Wu S, Zhang Y (2007) LOMETS: a local meta-threading-server for protein structure prediction. Nucleic Acids Res 35(10):3375–3382. https://doi.org/10.1093/nar/gkm251

Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER suite: protein structure and function prediction. Nat Methods 12(1):7–8. https://doi.org/10.1038/nmeth.3213

Yousef M, Abdelkader T, & ElBahnasy K (2017) A hybrid model to predict proteins tertiary structure. Paper presented at the 2017 12th international conference on computer engineering and systems (ICCES)

Zhu Z, Fan Y, Liu Y, Jiang T, Cao Y, Peng Y (2020) Prediction of antiviral drugs against African swine fever viruses based on protein-protein interaction analysis. PeerJ 8:e8855. https://doi.org/10.7717/peerj.8855

Chapter 18
Promises and Premises of Applications of Artificial Intelligence in Animal Health

Chandra Sekhar Mukhopadhyay and Yogendra Singh Jadoun

Abstract Artificial intelligence (AI) has become an integral part of our daily life. It is being used in almost every smart device or gadget that we are using every day. AI and its components, like machine learning and deep learning, find application in almost every sphere of life, including animal husbandry and one-health. The present chapter highlights the major applications of AI in animal sciences with special emphasis on its possible uses to enhance farm returns.

Keywords Artificial intelligence · Robotics · Mechatronics · Animal health

18.1 Introduction

It is well known to all of us that "Necessity is the mother of invention." Similarly, the need for a data economy is the basis for the evolution of artificial intelligence. Data economy refers to the pan-global digital ecosystem undertaking certain activities on data, like data compilation, organizing, sharing, and exchanging through a network of vendors to derive "value" from all gathered information. It also harnesses to determine the actual growth in terms of volume of data generation. In recent times, a similar word has surfaced quite meaningfully: "Digital Economy." It indicates the

Artificial intelligence is growing up fast, as are robots whose facial expressions can elicit empathy and make your mirror neurons quiver (Diane Ackerman)

C. S. Mukhopadhyay (✉)
College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India
e-mail: csmukhopadhyay@gadvasu.in

Y. S. Jadoun
Department of Veterinary and A.H. Extension Education, College of Veterinary Science, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India
e-mail: yogendrasingh@gadvasu.in

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_18

economy is based on digital technologies through the worldwide network of financial and commercial transactions and professional interactions. The underlying factor that has harnessed this digital vis-à-vis data economy is the information and communications technologies (ICT) empowered by several factors, like electronic gadgets, smart devices, wired and wireless internet connectivity, 5G, internet of things. In the next couple of years, till 2025, the next 5 years up to 2025, global data creation is projected to grow to more than 180 zettabytes. *The new data economy creates continuous combat for data proprietorship to derive returns from it. Big data is generated by the increase in data volume which helps manage a large quantity of data. Data can be analyzed with the help of data science*.

Now, we will shift to the concept of artificial intelligence and how training a machine can impact the same. We can teach machines to learn from data and derive a variety of acumens giving rise to artificial intelligence (AI). Artificial Intelligence is exhibited by a machine that simulates human and animal intellect. It involves intelligence agents, the independent entities that notice their environment and takes action that makes the best use of their probabilities of success at a given goal. Using logic allows computers to mimic human intelligence. It's a program that can act, reason, and sense.

18.2 Paradigms of Artificial Intelligence

Before diving deep into the various areas of applications of artificial intelligence in animal sciences, some information about artificial intelligence and its paradigm is necessary. Artificial intelligence (AI) is the intelligence exhibited by machines & its capability to mimic human behavior, viz., cognitive functions like problem-solving and learning. The term AI encompasses several independent vis-à-vis related disciplines of science. In Fig. 18.1, various paradigms, not necessarily limited to these only, have been depicted.

The quotation given below reflects the pervasiveness of AI and machine learning:

"Machine learning is a core, transformative way by which we're rethinking how we're doing everything. We are thoughtfully applying it across all our products, be it search, ads, YouTube, or Play. And we're in early days, but you will see us — in a systematic way — apply machine learning in all these areas" (Google CEO Sundar Pichai, 2016)

Artificial intelligence has now found its application in various domains. Some of the most important domains where machine learning has led the path for AI to rein are as follows:

- *Data preprocessing*: includes data wrangling and manipulation.
- *Supervised and unsupervised learning*: Regression, classification, clustering, etc. Unsupervised learning allows us to find previously unknown patterns in datasets that are neither classified nor labeled.

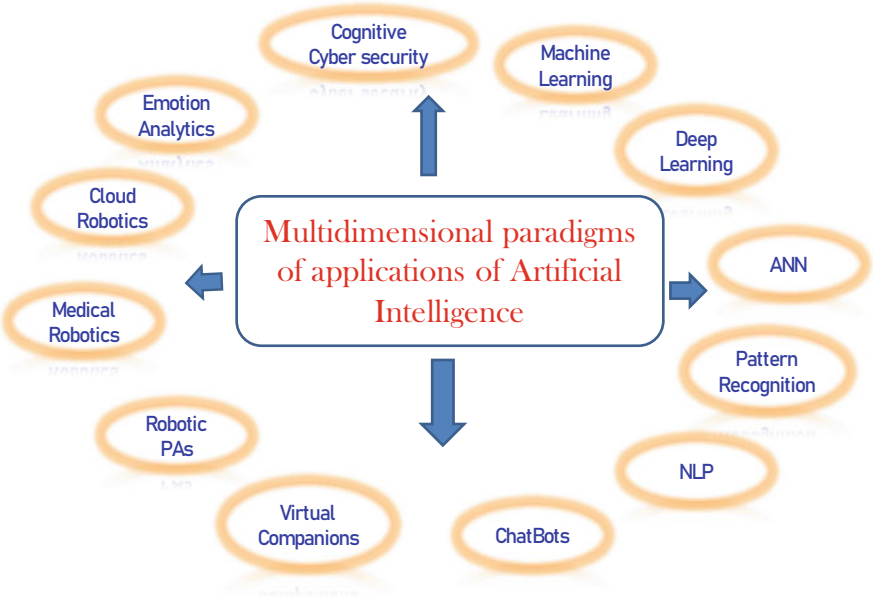

Fig. 18.1 Multidimensional applications of artificial intelligence

- *Time-series modeling*: A model is developed by sampling data points/values from equally spaced intervals.
- *Feature engineering*: It's predominantly used to select a certain feature from the given data.
- *Recommender system*: Perhaps, this system we are most intricately associated with. The e-commerce companies like Amazon, Azio, Flipkart use recommender systems to provide recommendations to their customers. Whenever we visit such e-commerce companies, the AI system notes the likes and dislikes of the customer and recommends the products according to the user's linking patterns (Fig. 18.2).

(a) *Text mining*: It's text analytics that is based on artificial intelligence technology for transforming unstructured text (from documents like online pages or Word files) into normalized, structured data. A machine learning–based method called natural language processing (NLP) is being used for text mining.
(b) *Ensemble learning*: It's an advanced aspect of ML where it's used to solve a particular computer intelligence problem.

These were the brief, but not exhaustive, areas where AI is very systematically used. Now, let us elaborate on the term machine learning.

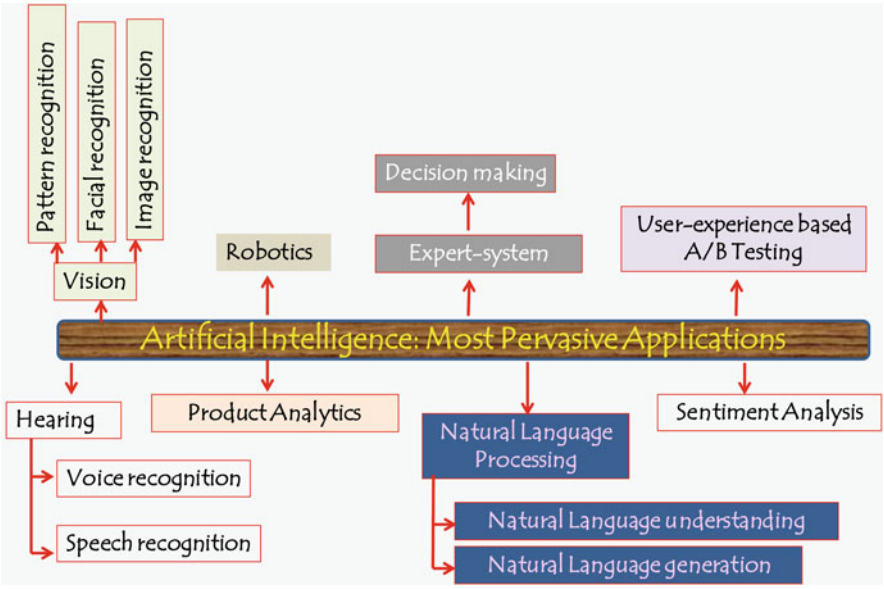

Fig. 18.2 Broad domains of usage of artificial intelligence in everyday life (Source: Android App: Artificial Intelligence: Applications in Animal Sciences)

18.3 Machine Learning

Machine learning (ML) is a subset of artificial intelligence. ML is an approach to achieving AI through systems or software models that can learn from an experience to find a pattern from a set of data. Thus, it provides a machine the ability to *learn automatically and improve from experience*. It involves teaching a computer to recognize data patterns, and ML creates rules rather than traditional programming.

When we study ML, we frequently come across three terms that sound quite similar: classification, clustering, and categorization. The differences between these three terms are explained below:

- *Classification:* The computer program learns from the data input provided to it and then uses this knowledge to categorize new annotations. Classification is used for forecasting discrete replies. It is used when we are training a model to expect qualitative goals.
- *Categorization:* A procedure to establish data into categories for its utmost effective and efficient usage. It makes pretext searches quicker and offers a better user understanding.
- *Clustering*: It is a technique of alliance a set of items in such a way that objects in the same group are most similar to each other than to those in other groups. Hence, it is the assembling of objects based on resemblance and dissimilarity among them.

18.4 Features of ML

ML uses data to notice patterns in a dataset and fine-tune program actions consequently. Pattern recognition can be defined as the classification of data based on the information previously gained or on statistical information taken out from the patterns.

(a) ML emphasizes the growth of computer programs that can teach themselves to grow and modify when exposed to novel data using a method called reinforcement learning. It uses outside feedback to teach the system to change its internal working to get improved next time.
(b) It allows the computer to search out hidden intuitions using iterative algorithms without being programmed. ML uses an algorithm that learns from preceding data to harvest reliable and repeatable decisions. The difference between machine learning and traditional programming has been adumbrated in Table 18.1.
(c) *It mechanizes logical model building using statistical and ML algorithms. That is patterns and relationships from data and expresses as mathematical equations*.

Table 18.1 Difference between traditional and machine learning approaches

SN	Traditional programming	ML
1	Data and programs are provided to the computer. It processes them and gives the output	The algorithm is applied to the given data and the yield. The outcome of the applied algorithm and calculations is a learning model, which aids machines to learn from the data
2	You code the behavior of the program	To learn from the data, we leave a lot of that to the machine
3	Traditional programming relies on hard-coded rules. The result is evaluated at the end. If the result is satisfactory, the program is deployed into production. However, if the result is not satisfactory, one will review the errors, change the programs, and evaluate it again. This iterative process is continued till one gets the expected result	In the ML approach, the decision rules are not hard-coded. The problem is solved by training the model with training data to derive or learn an algorithm that best represents the relationship between the input and the output. The trained model is evaluated against test data. Here onward, the best part is the same as traditional programming: If the result is satisfactory, the program is deployed into production. However, if the result is not satisfactory, one will review the errors, change the programs, and evaluate it again. This iterative process is continued till one gets the expected result

18.5 Applications of AI Techniques in Health Sciences

- *Medical/Veterinary image diagnosis*: USG, MRI images, skin cancer prediction.
- *Personalized medicin*e: based on individual genome analysis.
- Smart health records.
- Smart animal husbandry.
- *Mobile or computer apps* for instantaneous data recording and analysis.
- *Imaging Processing*: It converts an image to a digital format and performs some operation on it to induce or enhance an image or to extract some helpful information out of it. Example:
  - Automatic face-tagging by recognizing an image from a previously tagged photo.
  - Optical character recognition (OCR) is used to convert a printed document into a digital format (to digitize the text).
  - Another example is a self-driving car.
- *Health Care*: Identify disease; diagnosis; drug discovery; medical image processing and detection. Some companies have used ML and AI to revolutionize and transform the healthcare sector, viz. Google deep mind, Giner.io, health.
- *Robotics*: Robots are machines that can be used to do some jobs.
- Emotions of human beings can be read by the Humanoid robots.
- Industrial robots are used for assembling and manufacturing products.
- *Data Mining*: Method of analyzing hidden patterns in data. Example:
  - Anomaly detection (to detect credit card fraud to determine which transaction is anomalous from the usual pattern),
  - Association rules (market basket analysis: to detect which items are often bought together).
  - Grouping & predictions (to classify users based on their profiles).
- *Video Games*: to give predictions based on data. Example:
  - In Pokémon goes battle a lot of data taking into consideration to foresee the winner. The classifier of machine learning will calculate the result of the match based on the data.
- *Analysis of the Text*: It is an automated process of obtaining information from text. Example:
  - Spam filtering.
  - Sentiment analysis (to classify whether an opinion is positive, negative, or neutral. It detects public sentiment in Twitter feed or filters customer complaints).
  - Information extraction: extracting specific data addresses, keywords, or entities.

18.6 Artificial Intelligence (AI)–Based Solutions for Animal Sciences

In other sectors, artificial intelligence (AI) has at present established itself to be the main force for increasing competence and efficiency, and now is the time to enlarge its uses and application in the animal health sector. At present, we can already see the rise of artificial intelligence–based services and applications in Agriculture and allied sector operations, but still, these are in budding stages and need to travel the long journey to come into routine practices *(https://www.pashudhanpraharee.com)*. The important areas where AI has been used in animal sciences are discussed:

- *AI as an investigative tool:* Some veterinary practices and management procedures using a form of artificial intelligence for several years. Specifically, the majority relied on exclusive systems that entrench improved diagnostic tools and applications in their routine practice software. These programs recommend possible diagnostic testing substitutions to be next step treatments based on clinical signs and investigated results.
- *AI in complex diagnostic procedures*—AI is also used for multifaceted diagnostic instruments, devices, and procedures. The AI-imbued stethoscope can be used to assist in identifying heart-related ailments and also in various urine-related infections using AI applications and services. All these AI-based procedures help time-saving and escalate efficiency in the veterinary practice.
- *AI as imaging paraphernalia:* Various AI-based imaging applications and services help the vets practitioners and scientists to organize images for clinical clarification and increase practice efficiency meaningfully. This provides safety in building the diagnoses for veterinary practitioners.
- *Pet cameras and pet tracker system:* Pet animal cameras and pet animal trackers are a wealth house of pet data and information. These cameras and trackers screen virtually all of your pet's actions and movements daily, like drinking behavior, eating, walking, sleep patterns. This type of information opens up many opportunities for AI-based machine learning. Seeing the recording of a camera directs vigilant signs to check any alarming situation or discomfort to the animals.
- *Walking of dog:* From a natural language processing (NLP) system, dog-walking activities could take the benefit that carrying a review on a weekly basis to their customers about their pet's movements and activities. Owners of the dog could receive personalized information describing the duration, walk routes, encounters with other dogs and food, urination, and potty pauses.
- *Smart feeders and waterers:* AI-based smart feeders and waterers monitor the pet's eating and drinking movements, and habits centered on any erroneous happenings. For example, pet owners watch their animals for drinking habits that drink smaller amounts of water than normal, which may indicate regarding kidney problems that should be well-thought-out.
- *Communication barricades:* Removal of message blockade is vital, as other than human beings have imperfect kinds of verbal and non-verbal signs of

communication. These limitations can be sorted out by using specific signs and pictures that pet animals can identify easily.

- *AI in observations:* AI can be worthwhile to observe the individual happenings of each pet animal.
- *Sleep patterns and behavioral changes:* Consistent and troubled sleep patterns can aid in judgment the justifications with the movement of the brain, the efficient capability of overall body ease of individual or pets. Likewise, we all have sensitive stages that influence our actions and replies, also found among the animals. Each species has its mood swings like humans, which can help to generate AI-based interpretations and resolutions.
- *Thinking procedure and training:* We can identify more about the thinking procedure of pet animals, which is dissimilar from humans. It can help AI in the direction to draft the training programs for pets or animals. Disciplined, pet animals, which are the part of animal husbandry sector, need a varied amount of training. Visual and oral examination of needs can lead to advancement in training excellence by reducing complications.
- *Mechanization/automation:* Manual supervision of daily pursuits and training pets is labor as well as time intensive. However, the same task can be achieved in a smarter way through the applications of AI.
  - For example, the measurement of the thickness of fur and taking out wool sheep needs expert care, in number of times every winter, but by use of AI, it can be easily be done.
- *Analysis of data through AI:* Any type of documents relating to the weight of the animals, health issues, age, life expectancy, growing stage, reproduction period, puberty, bodily changes, hormonal changes, relationship with the human caretakers, and much more is all likely to be possible with the high-tech developments and interventions.
- *Research and development activities:* AI can take analysis and growth with the help of research scientists working in various institutes and come out with the solutions and AI-based developed technologies for proper recording and their solutions to key problems.
- *Predictions:* Humans are aware that animals have a great instinct for environmental fluctuations, and many times, it is found that they start moving from the place where disaster is going to happen to the safe places or surroundings. This can facilitate AI to alert human beings about the disaster or any kind of natural calamities or risks.
- *Disease monitoring:* The models are trained so that the healthy and diseased livestock/birds can be differentiated based on their physical appearance, nasal secretion, behavioral changes, etc. Besides, the same principle applies to wild animals, which can save those precious lives.
- *Drone-regulated flock management:* Management of large flocks of large and small ruminants, and even larger birds like emu, turkey is a hectic job. The loss of animals may ultimately lead to low economic return and certain legal issues if animals enter others' premises/farms. To avoid these unprecedented but

eventually possible incidences, drones are used to constantly monitor the movement of the flocks, and take and send pictures to the main database where the pictures are analyzed by computers, and an alarm is raised if needed. Besides, AI-driven drones can also monitor the behavior of the animals and study if there is an anomaly.

- *Use of AI in poultry farming/sector:* Development in the poultry farming sector has been rising, and to address the persisting challenges in commercial poultry farming, precise farming using digital gadgets and machinery holds promise. AI-driven mechatronics, robotics, etc. can directly contribute to smart decision making, operation of small to large farms/ poultry meat plants, minimizing labor cost, and subsequently improve the outcome and farm returns.
  - *Operational challenges of robotics:* Poultry meat processing involves dressing of carcass (following killing the birds adopting humane method), removal of inedible parts (like digestive and reproductive systems, head), separation of giblets, deboning, etc. The automated robots need to be thoroughly trained to recognize the viscera and the internal organs, vis-à-vis their shape and span. Sensors capture photos of these viscera and various organs so that the system can be trained through deep-learning models (which is a component of machine learning). Such rigorous training requires a lot of time to acquaint the robots with birds of various breeds, their visceral structures, and also their internal organs.
- *Detection of behavioral and nutritional diseases in birds*: Birds show certain odd behavior, like feather pecking, and attacking other birds, which ultimately leads to depraved health conditions for the birds. Intelligent systems have been developed to identify such birds in the flock so that they can be treated for any mineral deficiency or nutritional deficiency or removed such birds if required.
- *Applications of radiofrequency identification device (RFID) technology:* This technology uses digital data, which is encrypted in RFID tags and read by a reader thru radio waves. RFID systems comprise three (03) parts: an RFID reader, an RFID tag, and an antenna. An RFID tag comprises an integrated circuit and an antenna, which is used to transfer data to the RFID reader, which then translates the radio waves into a practical form of data. This information is transported via a communications interface to the main computer system, where the data is stored in a database and analyzed later on. *By placing RFID tags on agricultural product packages, farmers can analyze the health condition of the product, making it suitable for processing companies to synchronously add information on the tag, such as enterprise codes, batch processing, processing date, and weight of the packages.*

18.7 Conclusion

In the absence of artificial intelligence (AI), most veterinary practices and services cannot do a substantial impact with the gathered data they are gradually collecting and documenting on a routine basis. Data-driven AI services, applications, and tools could supplement speediness, lifesaving medication, and healthy customer communication in any veterinary practice.

Chapter 19
Computational Genomics Approaches for Livestock Improvement and Management

Kalpana Singh, Mir Asif Iquebal, Sarika Jaiswal, Anil Rai, and Dinesh Kumar

Abstract Livestock (domesticated animals such as cattle, buffalo) are important economic and genetic resources used in agriculture to provide labour, and commodities such as meat, eggs, milk, fur, leather, and wool. Due to their commercial importance, continuous efforts have been made by breeders and research community worldwide to increase resistance and reproductive potential for their improvement and better management. The advancement in next-generation sequencing approaches made it is possible to conduct various genome-wide studies at a lower cost with faster, efficient, and higher depth to analyse various diseases, diversity, breeds, resistance, reproduction, and developmental stages in livestock. In this chapter, we are providing a summary of various computational approaches utilized at the various level of study of genomics such as the genome assembly, molecular markers, epigenetics, and transcriptome. This chapter covers various computational approaches for genome assembly along with approaches related to genome assessment, structural annotations (localization of transposable elements, non-coding RNAs, RNAs, etc.), and functional annotation (identification and characterization of genes) post-assembly. Next, it covers computational approaches for identification and characterization of molecular markers (SNP, SSRs). This chapter also includes computational approaches for epigenomics to study DNA methylation (MeDIP-seq, BS-seq analyses) and histone modifications (ChIP-seq, ATAC-seq, DNase-seq analyses) along with the approaches for transcriptome (RNA-seq, sRNA-seq,

K. Singh
Division of Agricultural Bioinformatics, Earlier Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, PUSA, New Delhi, India

Department of Bioinformatics, College of Animal Biotechnology, Guru Angad Dev Veterinary And Animal Sciences University, Ludhiana, Punjab, India

M. A. Iquebal (✉) · S. Jaiswal · A. Rai · D. Kumar
Division of Agricultural Bioinformatics, Earlier Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, PUSA, New Delhi, India
e-mail: ma.iquebal@icar.gov.in; sarika@icar.gov.in; anil.rai@icar.gov.in; dinesh.kumar@icar.gov.in

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_19

Degradome-seq analyses). This chapter could be helpful to introduce students and researchers with various computational genomics approaches available for the improvement and management of livestock.

Keywords Animal science · Annotation · Epigenetics · Genome assembly · Molecular markers · Transcriptome

Livestock are domesticated animals such as cattle, buffalo, goat, sheep, horse, donkey, and pig. Livestock are important economic and genetic resources used in agriculture to provide labour, and commodities such as meat, eggs, milk, fur, leather, and wool. Due to their commercial importance, continuous efforts have been made by breeders and research community worldwide to increase resistant and reproductive potential for their improvement and better management. Due to advancement in next-generation sequencing approaches, it is possible to conduct various genome-wide studies such as genome assembly, transcriptome, and marker analysis, epigenetics at lower cost with faster, efficient, and higher depth to analyse various diseases, diversity, breeds, resistance, reproduction, and developmental stages in livestock. In this chapter, we are providing a summary of various computational approaches utilized at each level of genomics as follows.

19.1 Computational Approaches for Genome Assembly, Assessment, and Annotation

The aim of genome assembly is to determine, assemble, and organize the complete genome sequence of an organism into chromosomes, and, further, to annotate the protein-coding genes and other genetically important functional features such as transposable elements, microsatellites, and non-coding genes. During whole-genome sequencing, sequence reads of varying lengths are generated depending on the technique used; for example, smaller sequence reads of 50–300 base length are generated from second-generation sequencing, e.g. Illumina short reads; and long reads of 10,000–20,000 base length are generated from third-generation sequencing, e.g. PacBio long reads. These whole-genome reads are considerably shorter than the most genomes or even than the most genes. Therefore, putting the sequence reads into the correct order is crucial and called as genome assembly (Foxman 2014). A high-quality genome assembly is a prerequisite for the whole-genome analyses.

The continuous advancement in the next-generation sequencing technologies and increased computational power has reduced the cost of genome sequencing and made it high throughput greatly over the past decade. Yet, the current methods of genome sequencing still need improvement due to the nature of the genome sequence itself, and due to inherent problems in the sequencing methods. There are varying reasons; for example, the presence of repeats (variable number of repeated sequences), polyploidy (whole-genome duplications), aneuploidy (multiple

copies of chromosomes), polymorphism (different regions on homologous chromosomes), chimera (multiple genomes from the same sample of an organism), DNA sample from multiple organisms (to obtain enough DNA, i.e. requires more than one individual due to tiny size of organisms of the some species), variable localization of genes, variable localization of multiple copies of the same gene, sequencing errors, sample contamination, and expertise cause complication in genome assembly and result in ambiguities (Dominguez Del Angel et al. 2018). The sequencing errors caused due to sequencing machine may produce reads not covering the entire genome causing missed regions in the final assembly or due to varying base accuracy from different sequencing technologies.

There are three types of assemblers based on the approaches utilized. The first approach is a Greedy method (a rapid-assembly algorithm), followed by the overlap–layout–consensus (OLC) and Hamiltonian path method (mostly used for longer reads, >200 bp), and de Bruijn graph (DBG) and Eulerian path method (mostly used for short reads, <100 bp) (Li et al. 2012). Greedy first compares fragments in a pairwise fashion to identify overlapping sequences, followed by merging of the sequences with the best overlaps iteratively, till the last sequences with overlap. Some left unassembled reads are shown as gaps, which could be filled by paired-end sequencing. Many early extremely useful assemblers based on Greedy were Staden (Dear and Staden 1991), TIGR Assembler (Sutton et al. 1995), PHRAP (Green 1996), STROLL (Chen and Steven 2000), and CAP (Choudhuri 2014). The Phred–Phrap–Consed suite developed by Phil Green and Brent Ewing has been widely used. The OLC method calculates all the overlaps between the reads by pairwise comparisons, and prepares a directed graph using reads as nodes and overlaps as the edges. Then, it tries to find the Hamiltonian traversal path of the graph by combining the overlapping sequences in the nodes into the genome assembly to build the consensus sequence from the aligned overlapping reads. Some of the assemblers based on the OLC algorithm are SGA (most popular), CABOG (Celera Assembler) (Myers et al. 2000), Arachne (Batzoglou et al. 2002), Newbler (Margulies et al. 2005), MIRA (2023), Minimus (Sommer et al. 2007), and Edena (Hernandez et al. 2008). Overlap-based methods such as Greedy and OLC work well with sequence data with a finite number of reads to be assembled; in case of next-generation sequencing, hundreds of millions of sequence reads are generated, and then, the problem of scalability is solved by using the de Bruijn graph. The de Bruijn graph method breaks reads into shorter fragments, called k-mers (Compeau et al. 2011), aligned using $(k - 1)$ sequence overlaps, then builds a de Bruijn graph using all the k-mers. The size of k depends on sequence coverage, read length, etc. and should not be less than half of the actual read length. Finally, the genome sequences are inferred based on the de Bruijn graph. The DBG method also compensates for missing sequence reads. SPAdes (Bankevich et al. 2012a) is a popular assembler, which is based on the DBG method along with other as Euler-SR (Pevzner et al. 2004), Oases/Velvet (Zerbino et al. 2009), ALLPATH2 (Maccallum et al. 2009), ABySS (Simpson et al. 2009), and SOAPdenovo (Li et al. 2009a). In case of livestock genome, MaSuRCA (Zimin et al. 2013) works well for only short reads sequenced using Illumina and Solexa, etc. sequence analyser, as well as where both short and long reads are available. Canu (Koren et al.

2017) and FALCON (Gryan and Church 1994) work for only long reads sequenced using Roche 454 or PacBio sequence analyser. Platanus (Kajitani et al. 2014) and Platanus-allee (Kajitani et al. 2019) work well for highly heterozygous genome.

There are two ways for genome sequence assembly: reference-based mapping and assembly, and de novo assembly. In reference-based mapping and assembly, newly obtained resequenced reads are first mapped to already available reference genome through alignment and then assembled in proper order. Reference-based mapping and assembly is good for detection of single nucleotide variations and small InDels along with deletions and duplications by using coverage information and good for hiding the limitations of raw data. Bowtie (Langmead et al. 2009), Bowtie2, and BWA (Li and Durbin 2009) are ultrafast, more sensitive, and efficient short-read aligners based on mapping and assembly. In the absence of a reference genome sequence, de novo assembly is performed. De novo assembly is used to search unknown genes/transcripts, good for structural variations and detection of microsatellite. De novo assembly is a hierarchical process including the assembly of the sequence reads into contigs (segments with continuous sequence), the contigs into scaffolds (contiguous with N-filled gaps), and scaffolds into chromosomes. Due to the presence of gaps (which cannot be easily sequenced), many genome assemblies remain restricted to a scaffold level. Some of the scaffolds can be placed within a chromosome, while others may remain as unknown scaffolds (which could not be placed within any chromosome). Paired-end reads are better for de novo assembly in comparison to single-end reads.

Some of the available parameters for the assessment of the de novo genome assembly to decide whether the resulting assembly meets the standards are the available number of contigs and scaffolds and their sizes, and the fraction of reads that can be assembled. N50 value of contig and scaffold is one of the widely used parameters. To evaluate N50 contig value, first all contigs are sorted in a decreasing order of size; then, contigs are added from the longest contig, till the added size reaches at least half of the total size of all assembled contigs. The last contig added in this addition represents N50. Thus, an N50 contig is the minimum contig length required to cover 50% of the total length of the assembly. Other important parameters are L50 value (the number of contigs longer than N50), NG50 (minimum contig length required to cover 50% of the reference genome), NA50 (minimum length of aligned blocks required to cover 50% of the assembly), LG50 (number of contigs longer than NG50), LA50 (number of contigs longer than NA50), coverage (if $\leq$90% of the bases of genome have at least 5× read coverage, the genome is considered accurate), average contig length (longer than 5 kb), number of genes (assembly is considered better, if it identifies most of the known genes), number of gaps (decreases the quality of assembly), and genome fraction % (percentage of bases aligned to the reference genome) (Nagarajan and Pop 2013). QUAST (Gurevich et al. 2013), SQUAT (Yang et al. 2019), and GenomeQC (Manchanda et al. 2020) are used to measure varying assessment statistics of genome assembly.

For the assessment of contamination and completeness genome assembly, BUSCO: Benchmarking Universal Single-Copy Orthologue (Simão et al. 2015) provides evolutionarily sound measures of completeness and redundancy by

considering the presence of a predefined set of single-copy marker genes as a proxy for genome-wide completeness (Manni et al. 2021) in terms of expected gene content.

A robust and complete annotation is essential for a genome to be fully utilized by the scientific community, after a high-quality genome assembly. Genome annotation is the process of mapping genomic features (Harbola et al. 2022) such as pseudogenes, microsatellites, transposable elements, repeats, non-coding RNAs, SNPs, and regions with similarity to other genomes onto the genomic scaffolds along with protein-coding genes and their multiple mRNAs (structural annotation). Further, derivation of functional annotation of proteins or genes of the genome is important. Ironically, the rapid improvements in genome sequencing technology have reduced the accuracy of genome annotation, rather increasing it. The main challenges are the difficulty in the automated annotation of large and fragmented draft genomes (whose quality is uncertain), errors, and contamination in draft assemblies (Salzberg 2019). RNA-seq technology could provide a rapid solution to capture most of the expressed genes for any species due to its ability to determine exons and splice sites by aligning RNA-seq reads to a genome using TopHat (Trapnell et al. 2009), and then followed by the assembly of alignments (not the reads) into transcripts using cufflinks, which helps in the preparation of a reasonably good approximation (along with alternative isoforms) of the complete gene content of a species (Pertea et al. 2018).

Structural annotation comprises extraction of sequence features (sequence length, composition, GC content, transposable elements, and repeat identification) and the prediction of genes. The gene prediction is performed by identifying and localizing the open reading frames (ORFs), by identifying gene structures and coding regions, and by localizing the regulatory motifs. It is important to perform identification of repeats and transposable elements before gene prediction during genome annotation. For the identification of repeats (difficult due to their poorly conserved nature), first, repeat library of the genome is created by homology-based [MITE Digger (Yang 2013), MITE-hunter (Han and Wessler 2010), detectMITE (Ye et al. 2016), MITE Tracker (Crescente et al. 2018)] or by de novo [RepeatScout (Price et al. 2005), Recon (Bao and Eddy 2002), RepeatModeler2 (Flynn et al. 2020)] tools. Then, RepeatMasker (2019) uses BLAST (Altschul et al. 1990) to identify the homologs of repeat library in genome and masks the repeat regions (means denoting repeat regions as "N" in the genome such that the sequence alignment and prediction tools do not consider them for downstream processes). Transposable elements (TEs) are major players of genome structure, plasticity, genetic variations, and evolution due to their ability to move and accumulate in genomes (Dominguez Del Angel et al. 2018). TEs can affect gene expression, structure, and function due to their insertion in the vicinity of genes (Lisch 2013) and through epigenetic mechanisms (Slotkin and Martienssen 2007). REPET package (REPET 2019), one of the most used tools for large eukaryotic genomes and for efficiently detecting classified TEs (by TEdenovo), then annotates TEs into nested and degenerated copies (by TEannot). RepeatMasker is another of the most used tool for TE

identification and uses Dfam (Hubley et al. 2016) and Repbase (Bao et al. 2015), curated libraries of repeats.

For the prediction of genes, Augustus (Stanke and Morgenstern 2005) is the most extensively used tool and provides output in three files such as gff3, coding sequences (CDS), and protein sequences. Conrad (DeCaprio et al. 2007), CONTRAST (Gross et al. 2007), SNAP (Korf 2004), GlimmerHMM (Majoros et al. 2004), mGene (Schweikert et al. 2009), and GeneMarkS (Besemer et al. 2001) are other gene prediction tools. Prediction of tRNA genes could be performed using INFERNAL (Nawrocki and Eddy 2013), tRNAscan-SE (Chan and Lowe 2019), and RNAmmer (Christoffels and van Heusden 2019) along with other ncRNA (snoRNA, rRNA, tRNA, miRNA) gene prediction using RNA-code (Yuan and Sun 2013). After that, functional annotation of predicted genes could be performed to find their biochemical and biological functions using similarity search by BLAST and Interproscan (Quevillon et al. 2005); gene cluster prediction for secondary metabolites by antiSMASH (Medema et al. 2011); subcellular localization by WolfPSort (Horton et al. 2007); transmembrane domains by TMHMM (Krogh et al. 2001); gene ontology analysis by BLAST2GO (Conesa et al. 2005); and pathway analysis by KEGG (Erxleben and Grüning 2020).

The manual curation of assembled and annotated genome can be done by genome viewers such as Integrative Genome Viewer (IGV) (Thorvaldsdottir et al. 2013), GenomeView (Abeel et al. 2012), and Geneious (Kearse et al. 2012), which not only provide a comprehensive visualization but also possess features for modifying and updating predicted genes and their annotated features.

19.2 Computational Approaches for Molecular Marker Analyses

To study the genetic diversity within different breeds of buffalo, molecular markers such as single nucleotide polymorphism (SNP), InDels, and microsatellite/simple sequence repeat (SSR) markers are extracted due to high genetic variability among them. The high genetic diversity is an advantage for the success of breeding programs (Di Gaspero and Cattonaro 2010) as the characterization of the existing breeds is very useful to optimize the breeding programs, and to assist in the development of new and improved varieties (Muñoz-Espinoza et al. 2020).

SNP/InDel are the valuable source of genetic variability as these could derive the new genes or allelic variants selected by natural or artificial ways, when resultant phenotypes exhibit advantageous traits. SNP markers are widely used in GWAS, marker-assisted breeding, haplotype mapping, linkage disequilibrium, homozygosity mapping, candidate gene association study, population diversity study, and many other applications. GBS-seq (genotype by sequencing) or WGS (whole-genome sequencing) or reverse-seq data libraries first pass through quality check using KAT (Mapleson et al. 2017), ClinQC (Pandey et al. 2016), HTQC (Yang et al.

2013), BIGpre (Zhang et al. 2011), FastQC (Andrews 2010), and FASTX-toolkit (FASTX-Toolkit 2010) tools, followed by adapter removal using Fastx, Trimmomatic (USADELLAB 2016), Cutadapt/TrimGalore (Cutadapt 2023) tools, and removal of reads with Phred score of <20. Filtered reads then mapped to reference genome using BWT-based aligners such as Bowtie (Langmead et al. 2009), SOAP (Li et al. 2009a), SOAP2 (Li et al. 2009c), SOAP3 (Liu et al. 2012) and SOAP3-dp (Luo et al. 2013), and BWA (Li and Durbin 2009) or by using hash-based aligners such as RMAP (Smith et al. 2008), BFAST (Homer et al. 2009), MAQ (Li et al. 2008), and Novoalign and Stampy (Lunter and Goodson 2010). SNP/InDel markers are extracted from the aligned reads using QCall (Le and Durbin 2010), IMPUTE2 (Howie et al. 2009), realSFS (Nielsen et al. 2012), Beagle (Browning and Browning 2007), Atlas-SNP2 (Shen et al. 2010), SOAPsnp (Li et al. 2009b), MaCH (Li et al. 2010), Samtools/mpileup (Li et al. 2009d), GATK (McKenna et al. 2010), Playpus, SNVer (Wei et al. 2011), VarScan2 (Koboldt et al. 2012), and VarDict. According to Yao et al. (2020), Samtools/mpileup performed better in terms of sensitively and specificity followed by FreeBayes and GATK. SNPs could be of two types based on their localization in coding and non-coding regions. SNPs in coding regions are further of two types: synonymous (changing the amino acid during translation) and non-synonymous (missense causing a change to the same amino acid during translation or nonsense causing a change to the stop codon). There are varying machine learning tools for the prediction of the effects of SNP such as their effect of protein from moderate to drastic based on the missense mutation causing no effect to complete disruption of protein function by changing amino acid with similar/different size and physico-chemical properties. SnpEff (Cingolani et al. 2012), SIFT, Missense3D, SNAP2, SuSPect, PolyPhen, PredictSNP, variant effect predictor at Ensembl etc. are some of such tools. SSRs are the tandem repeats, and make up genomic repetitive regions along with the predominant interspersed repeats, remnants of transposable elements. SSRs are evolutionarily important due to their instability with higher rates of mutations than point mutations (Vieira et al. 2016). SSRs could be used to study diversity, gene flow, evolution, kinship among breeds, population structure analysis, to prepare linkage map, QTL mapping, marker-assisted selection, and other studies (Feng et al. 2016) due to their advantage being genetic codominant, multi-allelic (highly polymorphic), relatively abundant, widely dispersed across the genome, transferable between species, lesser required expertise, and easy automatically scored (Bhattarai et al. 2021). SSR with motifs of 03–05 nucleotides is more important. GBS-seq, WGS, and reverse-WGS sequencing reads first pass through quality check, followed by de novo assembly to retain the variability of SSR markers within each library using SOAPdenovo (Luo et al. 2012), IDBA (Peng et al. 2010), and SPAdes (Bankevich et al. 2012b). Later, SSR mining within de novo assembly is performed using MISA (Thiel et al. 2003), SSRLocator (Maia et al. 2008), TRF, SSRserver, SSRIT (Temnykh et al. 2001), MINE SSR (2005), FullSSR, Msatfinder, SSRFinder, GMATA, TROLL, SPUTINIK (Abajan 1994), and SciRoKo (Kofler et al. 2007). HipSTR program is used to extract polymorphic SSRs from different breeds. In silico primer designing is performed to design forward and backward

primers for SSRs using Primer3, Primer3plus, primer-blast etc. using parameters such as the annealing temperature of 59–61 °C, minimum GC of 40%, primer length of 18–25 bp, and 250-bp flanking sequences of SSRs for both primers.

A newly launched BuffGR (Khan et al. 2022) is an important genetic web resource specific to buffalo containing genome-wide SNP/InDels and SSRs along with their structural and functional annotation of five important breeds of buffalo including Murrah, an important Indian breed, and Mediterranean, an important breed worldwide.

19.3 Computational Approaches for Epigenetic Analyses

Epigenetics includes heritable changes in gene activity caused by mechanisms other than changes in DNA sequence. Epigenetic modifications are caused due to varying abiotic and biotic stresses, and result in change in gene expression. Epigenetic modifications involve DNA methylation, DNA acetylation, DNA–protein interactions, chromatin accessibility, histone modifications, and more, which cause heritable chemical or physical changes in chromatin (Ennis 2014) to activate or suppress gene activity.

DNA methylation involves addition of a methyl or hydroxymethyl group to nucleic acid bases of DNA sequence. However, the most studied DNA methylation is methylation of C5 position on cytosine bases, or m5c. Cytosine methylation (m5C) has a significant effect on regulation of spatial and temporal gene expression, chromatin remodelling (cause heterochromatin and suppression of transposable elements), and is critical for embryonic development and cellular differentiation processes (sp. stem cells). Perturbation in any of the above processes affected by methylation results in onset of disease. Enzymes responsible for the DNA methylation in mammalian cells are DNMT1 (DNA methyltransferase) isoforms, DNMT3L, DNMT3a, DNMT3b, and TET (ten–eleven translocation) (Guibert and Weber 2013). During DNA replication, DNMT1 isoforms methylate hemimethylated CpGs and maintain methylation patterns. In early developmental stages, DNMT3a and DNMT3b methylate unmethylated and methylated DNA as both perform as de novo methyltransferases. Similarly, DNMT3L also performs de novo methylation. TET enzymes perform methyl group oxidation by producing 5-hydroxymethylation as an intermediate. There are two methods for studying whole-genome or targeted DNA methylation, chromatin immunoprecipitation (ChIP), and bisulphite-based.

In the preparation of MeDIP-seq, 5-methylcytosine (5mC) monoclonal antibodies are used to purify methylated regions of the genome, which followed by sequencing (Neary and Carless 2020). After quality check, MeDIP-seq library reads are mapped to reference genome using Burrows-Wheeler Alignment Tool (BWA), Bowtie2, HISAT2, etc. alignment tools. MeQA is a quality control tool specific to MeDIP-seq. Initially, extraction of methylation was based on a peak calling method using MACS2 followed by DiffBind to extract differentially methylated regions. Later

over the time, MeDIP-seq–specific bioinformatics tools were developed to extract the methylated regions, followed by extraction of differentially methylated regions with the resolution of ~100 bp such as Batman, MEDIPS, BayMeth, MethRaFo, SIMD, DISMISS, MEDME, MeDUSA, DiffRep, and MeDEStrand. DMRs further pass through genomic localization, and structural and functional annotation analyses. coMET performs computation and visualization of functional genomic annotations and estimation of comethylation patterns of DMRs (Wilson and Beck 2016).

Bisulphite-based DNA methylation study methods (WGBS/BS-seq, RRBS-seq etc.) convert non-methylated cytosines to uracils with sodium bisulphite, which are read as thymine during PCR amplification, while m5C and hm5C are protected from conversion and read as cytosine. BS-seq is considered as the standard method in DNA methylation studies (Kurdyukov and Bullock 2016) and provide single base-pair resolution of DNA methylation. Further, contexts (CG, CHH, and CHG)-wise methylation could also be extracted using it. During analysis of BS-seq data, first quality-checked BS-seq libraries are mapped to reference genome. Then, the methylation estimation is performed using tools having wild card (allowing either Cs or Ts to map to Cs) such as BS Seeker (Guo et al. 2015), and tools aligning three letters (all Cs are converted to Ts) such as Bismark (Krueger and Andrews 2011). Later, differentially methylated regions are extracted using tools such as the bsseq package of R Bioconductor (Hansen et al. 2012), methylKit (Akalin et al. 2012), DMAP (Stockwell et al. 2014), MOABS (Sun et al. 2013), DSS-general (Park and Wu 2016), and metilene packages (Jühling et al. 2016), followed by functional annotation of DMRs (Shu et al. 2018).

Recently, epigenetic modifications of RNA such as m6A methylation were found and regulate RNA stability and localisation (Fu et al. 2014). They can be detected by MeRIP-Seq: methylated RNA immunoprecipitation sequencing (not possible to detect using BS-seq), which combines immunoprecipitation of modified sequences with RNA-seq analysis.

Other important epigenetic modification is histone modifications (covalent post-translational modifications), caused by addition or deletion of methyl, phosphorus, acetyl groups, ubiquitin, and/or sumo proteins to histone proteins, which results in change in gene expression by modifying chromatin structure or by recruiting histone modifiers. Various cellular processes such as transcription, gene silencing, cell cycle progression, apoptosis, differentiation, DNA replication, DNA repair/damage, nuclear import, and neuronal repression take place by transcriptional activation/inactivation, and chromosome packaging due to histone modifications. During histone acetylation, an acetyl group is added to histone H3 and H4 from acetyl coenzyme A by histone acetyltransferase (HAT) enzyme. On the contrary, during histone deacetylation, hydrolytic removal of acetyl group/s from lysine of histone takes place by histone deacetylases (HDACs) (Dokmanovic et al. 2007). During histone methylation, a transfer of one, two, or three methyl groups from *S*-adenosyl-L-methionine to lysine or arginine of histone takes place by histone methyltransferases (HMTs) (Greer and Shi 2012). On the contrary, during histone

demethylation, removal of methyl group from histones takes place by histone demethylases.

Histone modifications and DNA–protein interactions (with TFs: transcription factors, DNA binding proteins, etc.) are captured by the modified histones/DNA-bound proteins, and immunoprecipitated using a specific antibody. Then, to obtain the ChIP-seq libraries, the bounded DNA is coprecipitated, purified, and sequenced. ChIP sequencing captures transcription factor binding sites or histone modifications across the entire genome of any organism and helps in deciphering gene regulatory networks in combination with the transcriptome and DNA methylation analysis. Quality-checked ChIP-seq libraries are mapped with reference genome using sequence aligners such as MAQ (Li et al. 2008), Bowtie (Langmead et al. 2009), SOAP2 (Li et al. 2009e), BWA (Li and Durbin 2009), Novoalign and Stampy (Lunter and Goodson 2010), and Bowtie2 (Langmead and Salzberg 2012). Later, peak calling is performed to obtain the modified histone binding sites using tools such as MUSIC (Harmanci et al. 2014), MACS (Zhang et al. 2008), MACS2, ZINBA (Rashid et al. 2011), GEM (Guo et al. 2012), SICER (Zang et al. 2009), BCP (Xing et al. 2012), and F-Seq (Boyle et al. 2008a). Differential binding sites are extracted using tools such as DiffBind, DEseq2, ChIPDiff (Xu et al. 2008), and ODIN (Allhoff et al. 2014). Later, genomic localization (using HOMER by Heinz et al. 2010) and functional annotation are performed for differential binding sites using GREAT (McLean et al. 2010) for pathway analysis along with GO analysis using ChipPeakAnno (Zhu et al. 2010) and Blast2GO.

To study the RNA–protein interactions, immunoprecipitation-based RIP-Chip (RNA immunoprecipitation coupled with microarray analyses) was developed (Khalil and Rinn 2011). Recently, a crosslinking and immunoprecipitation (CLIP) protocol was developed by Darnell (2010) to overcome the shortcomings of RIP-Chip. Further, for the genome-scale identification of direct RNA targets (Riley and Steitz 2013), high-throughput sequencing of crosslinked RNA fragments (HITS-CLIP) (Licatalosi et al. 2008) is performed.

For the determination of whole-genome chromatin accessibility, various techniques such as FAIRE-seq (Formaldehyde-Assisted Isolation of Regulatory Elements sequencing by Giresi et al. 2007), DNase-seq (genome-wide map of DNase I hypersensitive sites, followed by sequencing by Boyle et al. 2008b), and ATAC-seq (an assay for transposase-accessible chromatin with sequencing by Buenrostro et al. 2015) are available. Chromatin accessibility could be utilized to study nucleosome mapping, transcription factor binding sites, and identification of novel enhancer, explore disease-relevant regulatory mechanisms, cell type–specific regulation, evolution, discovery biomarkers, and comparative epigenomics. ATAC-seq is a popular and better method, which sequences the regions of open chromatin to know how chromatin packaging takes place and how other factors affect gene expression without a prior knowledge of regulatory elements. ATAC-seq helps to find the chromatin accessibility, transcription factor binding sites, and gene regulation during complex diseases such as cancer, embryonic development, and T-cell activation. In ATAC-seq, genomic DNA is fragmented by insertion of a highly active transposase, Tn5 at open chromatin sites, and adding sequencing primers, called tagmentation.

Peak calling is performed after aligned ATAC-seq reads, followed by extraction of differential peaks if needed and functional annotation of peaks, similar to ChIP-seq data analysis.

19.4 Computational Approaches for Transcriptome Analyses

Transcriptome is the collection of all RNA molecules of an entire organism or a specific cell type including coding (mRNA) and non-coding (rRNA, tRNA, lncRNA, pri-miRNA, and others) RNAs (Botchkareva 2017). Non-coding RNAs may play regulatory role to alter gene expression at a genome level, transcriptome level, and post-translational level by regulation of DNA replication, RNA splicing, transcriptional regulation, and post-translational level. Transcriptome analysis may apply to understand that how genetic variations and different environmental factors (biotic and abiotic) contribute to altered expression, whole-transcriptome analysis are crucial. The advancement and reduced cost of next-generation sequencing techniques made such studies involving whole-transcriptome analyses easier, which are providing great insights into understanding various diseases due to genetic variations, and differential gene expression during various environmental stresses, and to uncover novel RNA species (Hrdlickova et al. 2017). It made possible to find out the biological pathways and molecular mechanisms involved in regulation of cell differentiation, development, and disease progression causing altered expression.

Nowadays, high-throughput RNA sequencing (RNA-seq) has replaced previous approaches such as microarray (Bayega et al. 2018). RNA-seq, first, extracts bulk RNA from sample and copies it into more stable double-stranded cDNA followed by sequencing of cDNA (Zhang et al. 2018). RNA-seq is capable of detection of all mRNA transcripts along with the regulatory siRNA and lncRNA transcripts both in low (as little 50 pg, even possible to analyse single-call scRNA-seq) and high abundance (in nano to micrograms to analyse bulk RNA-seq), along with identification of alternative splicing, novel transcripts, single-nucleotide polymorphism (SNP), and fusion genes. By using oligonucleotide probes binding with poly-A tails of mRNAs, the mRNA molecules can be separated.

For the gene expression study, RNA-seq libraries (full-length cDNA sequencing) for mRNA extraction, first, pass through quality assessment using tools such as FastQC (Andrews 2010), HTQC, RSeQC, FASTX-Toolkit (2010), RobiNA package (Lohse et al. 2012), and FaQCs, followed by adapter removal using tools such as Fastx, Cutadapt, Trimmomatic, TrimGalore, and removal of reads with Phred score of <20. Filtered reads then assembled de novo using assemblers such as SOAPdenovo-Trans, Trans-ABySS, Trinity (Grabherr et al. 2011), Bridger (Chang et al. 2015), rnaSPAdes (used for scRNA-seq) (Bushmanova et al. 2019), and Velvet/Oases (Zerbino and Birney 2008), and provide transcripts with alternative splicing and different expression levels. In case of availability of reference genome,

reference-based assemblers use aligned reads. However, reference-based assemblers fail to extract the structural alterations such as alternative splicing of mRNA transcripts. De novo assembly should be performed even with the availability of reference genome to recover transcripts transcribing from the missing segments of the reference genome assembly. Some of the splice junction–aware aligners are GMAP (Wu and Watanabe 2005), MapSplice (Wang et al. 2010a), GSNAP (Wu and Nacu 2010), RUM (Grant et al. 2011), TopHat (Trapnell et al. 2009), STAR (Dobin et al. 2013), Subread (Liao et al. 2013), HISAT, and HISAT2 (Kim et al. 2015), which find de novo junctions or use splice junctions from reference, if provided as in case of STAR aligner. Reference-based assemblers such as RSEM (Merino et al. 2019), Alexa-seq (Griffith et al. 2010), Cufflinks (Trapnell et al. 2010), StringTie (Pertea et al. 2015), FluxCapacitor (Montgomery et al. 2010), and MISO (Katz et al. 2010) assemble transcripts from aligned reads and estimate their relative abundance depending on how many reads support each one [count based by Sailfish (Patro et al. 2014), cuffquant, FeatureCounts (Liao et al. 2014), maxcounts (Finotello et al. 2014), FIXSEQ (Hashimoto et al. 2014), Rcount (Schmid and Grossniklaus 2015), htseq-count (Anders et al. 2015), and kallisto (Bray et al. 2016), but FRPK based by cufflinks]. After estimation of quantitative abundance of each transcript, differential gene expression is estimated by normalizing, modelling, and statistically analysing transcripts from two treatment comparisons using tools such as baySeq (Hardcastle and Kelly 2010), DESeq (Anders and Huber 2010), DEGseq (Wang et al. 2010b), edgeR (Robinson et al. 2010), PoissonSeq, DEseq2, Cuffdiff (Trapnell et al. 2013), Ballgown (Frazee et al. 2015), and Sleuth (Pimentel et al. 2017). Further, transcripts are annotated to genes using annotation file of reference genome if available, or by mapping against nr data of NCBI. TransDecoder is used for calling novel ORFs in RNA-seq assemblies from Trinity, Cufflinks, and other de novo assemblers. Further, pathway and gene set analysis is performed using SeqGSEA (Wang and Cairns 2014), PathwaySeq, BioCyc, ToPASeq, GAGE (Luo et al. 2009), PathView (Luo et al. 2013), GoSeq (gene ontology analysis for RNA-seq) (Young et al. 2010), GSAASeqSP (Xiong et al. 2014), FunRich, RNA-Enrich, GOexpress, and GOseq. Coexpression analysis among genes is performed using Cytoscape (Shannon et al. 2003), petal, BRANE Clust, GeneNetWeaver, WGCNA, and Pigengene. Genome browsers such as IGB (Integrated Genome Browser) (Nicol et al. 2009), Savant Genome Browser 2 (Fiume et al. 2012), and IGV (Thorvaldsdottir et al. 2013) are used to display RNA-seq alignment files and are capable of representing unique features of RNA-seq data such as strands, exons, introns, splice sites, and exon junction read counts (Griffith et al. 2015).

sRNA-seq (small RNA sequencing) is used to extract small RNA below 30 nt, such as miRNA (microRNA), piRNAs, and endosiRNAs. miRNAs are captured efficiently by direct ligation with adapters without any additional processing of RNA before ligation, followed by RT, PCR and sequencing. However, this method causes substantial biases due to influence of sequence on ligation, mitigated by adding degenerate random nucleotides to the ligation ends of adapters (Hrdlickova et al. 2017). miRNAs modulate gene expression at post-transcriptional level by targeting mRNA, lncRNA etc. by cleaving or sequestering them, and function as biomarkers

for various diseases. During sRNA-seq analysis for miRNA, first, after quality check and adapter trimming, clean reads are mapped with Rfam (a curated database for RNA families of structural RNAs including non-coding RNA genes and *cis*-regulatory elements) (Kalvari et al. 2020) and TREP (a curated database of transposable elements (TEs) (Wicker et al. 2007) to remove rRNAs and TEs, respectively. First, conserved miRNAs are extracted by mapping the remaining reads with miRBase (Kozomara and Griffiths-Jones 2010), miRNA database using BLAST or miReader (Jha and Shankar 2013) allowing six nucleotides at 3′ end and three nucleotides at 5′ end considering imperfect DICER processing. RumimiR (Bourdon et al. 2019) is a miRNA database specific to ruminants. Further, unmapped reads with point mutations or gone through RNA editing are then loosely mapped with miRNA precursor sequences. Still, unmapped reads are used to find novel miRNAs by aligning with reference genome and pass through two approaches such as RNA folding method and star strand expression method to find novel miRNAs. In RNA folding method–based tools such as deepMiRGene (Park et al. 2016), miRNAFold (Tempel and Tahi 2012), RNAmicro (Hertel and Stadler 2006), and miRNAminer (Artzi et al. 2008), first, aligning sequences with genome were extracted along with ~100-nt flanking sequences of both ends and then pass through RNA folding tools. Folded sequences with minimum free energy <25 kcal/mol and having main sequence lying in one arm of hairpin are filtered as putative miRNA precursors. After further, trimming precursor sequencings are refolded to ensure stability. Resulting precursors with mature miRNA sequence are considered novel miRNA. In star strand expression method–based tools such as miRdeep (Yang and Li 2011), aligned reads with genome are used to identify novel miRNAs based on higher expression of mature miRNA than the star strand and loop sequences due to DICER processing. Finally, read counts for each (known and novel) miRNAs are normalized by the total number of mapped miRNAs to measure the abundance of each miRNA. After calculation of abundance of each miRNA, differential expression of miRNAs could be calculated using tools used for gene expression analysis from RNA-seq data.

miRNA negatively control translation of mRNA or cause mRNA degradation by cleavage or sequestering them. Many in silico methods for miRNA target prediction are based on the fact that target binding site of miRNAs is conserved through different species. Therefore, many target prediction tools are based on conservation status such as EMBL (Stark et al. 2005), PicTar (Krek et al. 2005). PITA Top (Kertesz et al. 2007), EIMMo (Gaidatzis et al. 2007), mirWIP (Hammell et al. 2008), miRBase Targets (Griffiths-Jones et al. 2008), miRanda (Betel et al. 2010), DIANA-microT (Reczko et al. 2012), TargetScan (Nam et al. 2014), and DeepTarget (Lee et al. 2016), while some tools are independent of conservation status such as PITA All (Kertesz et al. 2007), RNA22 (Miranda et al. 2006), and TargetScan All (Nam et al. 2014). For the validation of mRNA targets of miRNAs and identification of miRNA cleavage sites, Degradome sequencing (Degradome-Seq) or parallel analysis of RNA ends (PARE) is used. It is a modification of 5′-Rapid Amplification of cDNAEnds (RACE) using high-throughput deep sequencing methods.

Degradome-seq could find the known and novel miRNA targets. StarBase is a database for miRNA cleavage sites obtained from Degradome-seq.

Circular RNA generated by back-splicing forms a covalently closed continuous loop, lacks polyadenylated tails, and localizes in cytoplasm. In circRNA-seq, first linear RNAs are digested using exonuclease R, followed by regular RNA-seq methods involving fragmentation and RT-PCR. RNA-seq libraries, first, were quality-checked and then aligned with the reference genome using Bowtie, Bowtie2, STAR, BWA, BLAT (Kent 2002), and TopHat-Fusion (Salzberg 2019). There are two groups of prediction tools based on the type of approaches used by them. One group of tools such as KNIFE (Szabo et al. 2015), PTESFinder (Izuogu et al. 2016), NCLScan (Chuang et al. 2016), and CIRCexplorer (Zhang et al. 2014) is based on gene annotation, termed as pseudo-reference–based or candidate-based approach. Another group of tools such as circRNA_finder (Westholm et al. 2014), CIRI (Gao et al. 2015), find_circ (Memczak et al. 2013), CIRCexplorer, DCC (Cheng et al. 2016), MapSplice (Wang et al. 2010), UROBORUS (Song et al. 2016), and Segemehl (Hoffmann et al. 2014) are based on fragment-based approach or segmented read approach by identifying back-splicing junctions from the alignment of multiple-split reads to the genome. Further, annotation of origin of predicted circRNAs is performed using annotation file of reference genome. In case of different treatment conditions, Ribonuclease R–treated reads are used to extract differential expression of circRNAs. circBase (Glazar et al. 2014) is a database for bona fide circRNAs including their annotation.

lncRNAs are ≥200-nucleotide-long non-coding RNA containing polyA tails, with variable splicing, and regulate gene expression at epigenetics, transcription, and post-transcription level. lncRNAs are used in genetic improvement of species, to cure various diseases or as causal of disease occurrence, development, and diagnosis. lncRNAs have strong tissue specificity and are low in expression. Based on localization and orientation in genome, lncRNAs are categorised as sense, antisense, intergenic, and intronic lncRNAs. RNA-seq libraries were prepared after quality assessment of total RNA, library preparation, and sequencing, and used for the extraction of lncRNA. RNA-seq data pass through quality check, adaptor removal, mapping with reference genome, transcript assembly, and transcript count similar to RNA-seq data analysis for mRNAs. To identify lncRNAs from the transcripts with ≥200 nucleotide and number of exons ≥2, coding potential evaluation of all transcripts is performed by Coding Potential Calculator (CPC) analysis by CPC (Kong et al. 2007) and CPC2 (Kang et al. 2017) tools (transcripts with coding potential <-1 are considered as non-coding), Coding-Non-Coding Index (CNCI) analysis using CNCI (Sun et al. 2013), CNIT (Guo et al. 2019) tools, PFAM protein domain analysis (Mistry et al. 2021), and PhyloCSF (Phylogenetic Codon Substitution Frequencies) analysis (Lin et al. 2011). Further, differential expression analysis similar to mRNA is performed for filtered lncRNAs. Functional enrichment of lncRNAs can be predicted using WAFNRLTG (Li et al. 2022), lncRNATargets (Hu and Sun 2016), lncTar (Li et al. 2015), and LncRRIsearch (Fukunaga et al. 2019) to find mRNA targets of lncRNAs. Interaction of lncRNAs with miRNAs could be predicted using LncRRIsearch and GCLMI (Huang et al. 2019). Further,

ncRNA–protein interaction could be predicted using RPI-SAN (Yi et al. 2018) and RPITER (Peng et al. 2019). Finally, Cytoscape could be used to visualize all interacting molecules. DIANA-LncBase (Karagkouni et al. 2019) is a database of experimentally supported lncRNA targets of miRNAs, and LncAtlas database (Mas-Ponte et al. 2017) includes subcellular localization dataset for lncRNAs.

References

Abajan C (1994) SPUTINIK. http://espressosoftware.com/sputnik/index.h

Abeel T, Van Parys T, Saeys Y, Galagan J, Van de Peer Y (2012) GenomeView: a next-generation genome browser. Nucleic Acids Res 40(2):e12

Akalin A, Kormaksson M, Li S et al (2012) methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol 13:R87. https://doi.org/10.1186/gb-2012-13-10-r87

Allhoff M, Seré K, Chauvistré H, Lin Q, Zenke M, Costa IG (2014) Detecting differential peaks in ChIP-seq signals with ODIN. Bioinformatics 30(24):3467–3475. https://doi.org/10.1093/bioinformatics/btu722

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106

Anders S, Pyl PT, Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31(2):166–169. https://doi.org/10.1093/bioinformatics/btu638

Andrews, S. (2010) FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Artzi S, Kiezun A, Shomron N (2008) miRNAminer: a tool for homologous microRNA gene search. BMC Bioinform 9(1):39. https://doi.org/10.1186/1471-2105-9-39

Bankevich A, Nurk S, Antipov D et al (2012a) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5):455–477. https://doi.org/10.1089/cmb.2012.0021

Bankevich A et al (2012b) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5):455–477. https://doi.org/10.1089/cmb.2012.0021

Bao Z, Eddy SR (2002) Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res 12:1269–1276

Bao W, Kojima KK, Kohany O (2015) Repbase update, a database of repetitive elements in eukaryotic genomes. Mob DNA 6:11

Batzoglou S, Jaffe DB, Stanley K, Butler J, Gnerre S, Mauceli E, Berger B, Mesirov JP, Lander ES (2002) ARACHNE: a whole-genome shotgun assembler. Genome Res 12:177–189

Bayega A, Fahiminiya S, Oikonomopoulos S, Ragoussis J (2018) Current and future methods for mRNA analysis: a drive toward single molecule sequencing. Methods Mol Biol 1783:209–241. https://doi.org/10.1007/978-1-4939-7834-2_11

Besemer J, Lomsadze A, Borodovsky M (2001) GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 29(12):2607–2618. https://doi.org/10.1093/nar/29.12.2607

Betel D, Koppal A, Agius P, Sander C, Leslie C (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol 11:R90. https://doi.org/10.1186/gb-2010-11-8-r90

Bhattarai G, Shi A, Kandel DR et al (2021) Genome-wide simple sequence repeats (SSR) markers discovered from whole-genome sequence comparisons of multiple spinach accessions. Sci Rep 11:9999. https://doi.org/10.1038/s41598-021-89473-0

Botchkareva NV (2017) The molecular revolution in cutaneous biology: noncoding RNAs: new molecular players in dermatology and cutaneous biology. J Invest Dermatol 137(5):e105–e111. https://doi.org/10.1016/j.jid.2017.02.001

Bourdon C, Bardou P, Aujean E, Le Guillou S, Tosser-Klopp G, Le Provost F (2019) RumimiR: a detailed microRNA database focused on ruminant species. Database (Oxford) 2019:baz099. https://doi.org/10.1093/database/baz099

Boyle AP et al (2008a) High-resolution mapping and characterization of open chromatin across the genome. Cell 132:311–322

Boyle AP, Guinney J, Crawford GE, Furey TS (2008b) F-Seq: a feature density estimator for high-throughput sequence tags. Bioinformatics (Oxford, England) 24:2537–2538

Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34(5):525–527. https://doi.org/10.1038/nbt.3519

Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet 81:1084–1097

Buenrostro J, Wu B, Chang H, Greenleaf W (2015) ATAC-seq: a method for assaying chromatin accessibility genome-wide. Curr Protoc Mol Biol 109:21.29.1–21.29-9

Bushmanova E, Antipov D, Lapidus A, Prjibelski AD (2019) rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. GigaScience 8(9):giz100. https://doi.org/10.1093/gigascience/giz100

Chan PP, Lowe TM (2019) tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol 1962:1–14. https://doi.org/10.1007/978-1-4939-9173-0_1

Chang Z, Li G, Liu J, Zhang Y, Ashby C, Liu D et al (2015) Bridger: a new framework for de novo transcriptome assembly using RNA-seq data. Genome Biol 16(1):30. https://doi.org/10.1186/s13059-015-0596-2

Chen T, Steven SS (2000) A case study in genome-level fragment assembly. Bioinformatics 16(6):494–500. https://doi.org/10.1093/bioinformatics/16.6.494

Cheng J, Metge F, Dieterich C (2016) Specific identification and quantification of circular RNAs from sequencing data. Bioinformatics 32(7):1094–1096

Choudhuri S (2014) Additional bioinformatic analyses involving nucleic-acid sequences. Bioinformatics for beginners. Academic Press, New York, pp 157–181. https://doi.org/10.1016/B978-0-12-410471-6.00007-4

Christoffels A, van Heusden P (2019) Genome annotation: perspective from bacterial genomes. In: Encyclopedia of bioinformatics and computational biology. Academic Press, New York, pp 152–156. https://doi.org/10.1016/B978-0-12-809633-8.20092-7

Chuang TJ, Wu CS, Chen CY, Hung LY, Chiang TW, Yang MY (2016) NCLscan: accurate identification of non-co-linear transcripts (fusion, trans-splicing and circular RNA) with a good balance between sensitivity and precision. Nucleic Acids Res 44(3):e29

Cingolani P, Platts AE, Wang LL, Coon M, Nguyen T, Wang L, Land S, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly 6:80–92. https://doi.org/10.4161/fly.19695

Compeau PE, Pevzner PA, Tesler G (2011) How to apply de Bruijn graphs to genome assembly. Nat Biotechnol 29(11):987–991

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18):3674–3676. https://doi.org/10.1093/bioinformatics/bti610

Crescente J, Zavallo D, Helguera M et al (2018) MITE tracker: an accurate approach to identify miniature inverted-repeat transposable elements in large genomes. BMC Bioinform 19:348. https://doi.org/10.1186/s12859-018-2376-y

Cutadapt (2023). http://cutadapt.readthedocs.io/en/stable/index.html

Darnell RB (2010) HITS-CLIP: panoramic views of protein-RNA regulation in living cells. Wiley Interdiscip Rev RNA 1:266–286

Dear S, Staden R (1991) A sequence assembly and editing program for efficient management of large projects. Nucleic Acids Res 19:3907–3911

DeCaprio D, Vinson JP, Pearson MD, Montgomery P, Doherty M et al (2007) Conrad: Gene prediction using conditional random fields. Genome Res 17:1389–1398

Di Gaspero G, Cattonaro F (2010) Application of genomics to grapevine improvement. Aust J Grape Wine R 16:122–130

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21

Dokmanovic M, Clarke C, Marks PA (2007) Histone deacetylase inhibitors: overview and perspectives. Mol Cancer Res 5(10):981–989

Dominguez Del Angel V, Hjerde E, Sterck L et al (2018) Ten steps to get started in genome assembly and annotation. F1000Res 7:ELIXIR-148. https://doi.org/10.12688/f1000research.13598.1

Ennis C (2014) Epigenetics 101: a beginner's guide to explaining everything. The Guardian newspaper

Erxleben A, Grüning B (2020) Genome annotation (galaxy training materials). https://training.galaxyproject.org/training-material/topics/genome-annotation/tutorials/genome-annotation/tutorial.html. Accessed 24 May 2022

FASTX-Toolkit (2010). http://hannonlab.cshl.edu/fastx_toolkit

Feng S, He R et al (2016) Development of SSR markers and assessment of genetic diversity in medicinal *Chrysanthemum morifolium* cultivars. Front Genet 7:00113. https://doi.org/10.3389/fgene.2016.00113

Finotello F, Lavezzo E, Bianco L, Barzon L, Mazzon P, Fontana P, Toppo S, Di Camillo B (2014) Reducing bias in RNA sequencing data: a novel approach to compute counts. BMC Bioinform 15(Suppl 1):S7. https://doi.org/10.1186/1471-2105-15-s1-s7

Fiume M, Smith EJM, Dario AB et al (2012) Savant Genome Browser 2: visualization and analysis for population-scale genomics. Nucleic Acids Res 40(W1):W615–W621. https://doi.org/10.1093/nar/gks427

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF (2020) RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A 117(17):9451–9457

Foxman B (2014) A primer of epidemiologic study designs. Bioinformatics for beginners. Academic Press, New York, pp 41–52. https://doi.org/10.1016/B978-0-12-374133-2.00004-6

Frazee AC, Pertea G, Jaffe AE, Langmead B, Salzberg SL, Leek JT (2015) Ballgown bridges the gap between transcriptome assembly and expression analysis. Nat Biotechnol 33(3):243–246. https://doi.org/10.1038/nbt.3172

Fu Y et al (2014) Gene expression regulation mediated through reversible m^6A RNA methylation. Nat Rev Gen 15:293–306

Fukunaga T, Iwakiri J, Ono Y, Hamada M (2019) LncRRIsearch: a web server for lncRNA–RNA interaction prediction integrated with tissue-specific expression and subcellular localization data. Front Genet 10:462. https://doi.org/10.3389/fgene.2019.00462

Gaidatzis D, van Nimwegen E, Hausser J, Zavolan M (2007) Inference of miRNA targets using evolutionary conservation and pathway analysis. BMC Bioinform 8:69. https://doi.org/10.1186/1471-2105-8-69

Gao Y, Wang J, Zhao F (2015) CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. Genome Biol 16:4

Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD (2007) FAIRE (formaldehyde-assisted isolation of regulatory elements) isolates active regulatory elements from human chromatin. Genome Res 17(6):877–885. https://doi.org/10.1101/gr.5533506

Glazar P, Papavasileiou P, Rajewsky N (2014) circBase: a database for circular RNAs. RNA 20:1666–1670

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29(7): 644–652. https://doi.org/10.1038/nbt.1883

Grant GR, Farkas MH, Pizarro AD, Lahens NF, Schug J, Brunk BP, Stoeckert CJ, Hogenesch JB, Pierce EA (2011) Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). Bioinformatics 27:2518–2528

Green P (1996) Documentation for Phrap. University of Washington, Seattle. http://bozeman.mbt.washington.edu

Greer EL, Shi Y (2012) Histone methylation: a dynamic mark in health, disease and inheritance. Nat Rev Genet 13:343–357

Griffith M, Griffith OL, Mwenifumbo J, Goya R, Morrissy AS, Morin RD, Corbett R, Tang MJ, Hou YC, Pugh TJ et al (2010) Alternative expression analysis by RNA sequencing. Nat Methods 7:843–847

Griffith M, Walker JR, Spies NC, Ainscough BJ, Griffith OL (2015) Informatics for RNA-seq: a web resource for analysis on the cloud. PLoS Comput Biol 11(8):e1004393

Griffiths-Jones S, Saini HK, Dongen SV, Enright AJ (2008) miRBase: tools for microRNA genomics. Nucleic Acids Res 36:D154–D158

Gross SS, Do CB, Sirota M, Batzoglou S (2007) CONTRAST: a discriminative, phylogeny-free approach to multiple informant de novo gene prediction. Genome Biol 8(12):R269

Gryan G, Church GM (1994) Falcon: fast assemblies of large contigs. http://arep.med.harvard.edu/labgc/falcon.html. Harvard Medical School, Boston

Guibert S, Weber M (2013) Functions of DNA methylation and hydroxymethylation in mammalian development. Curr Top Dev Biol 104:47–83

Guo Y, Mahony S, Gifford DK (2012) High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints. PLoS Comput Biol 8: e1002638

Guo X, Gao L, Wang Y, Chiu DK, Wang T, Deng Y (2015) Advances in long noncoding RNAs: identification, structure prediction and function annotation. Brief Funct Genom 15(1):38–46

Guo JC, Fang SS et al (2019) CNIT: a fast and accurate web tool for identifying protein-coding and long non-coding transcripts based on intrinsic sequence composition. Nucleic Acids Res 47 (W1):W516–W522. https://doi.org/10.1093/nar/gkz400

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075

Hammell M, Long D, Zhang L, Lee A, Carmack CS, Han M, Ding Y, Ambros V (2008) mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. Nat Methods 9:813–819

Han Y, Wessler SR (2010) Mite-hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. Nucleic Acids Res 38(22):199

Hansen KD, Langmead B, Irizarry RA (2012) BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome Biol 13:R83

Harbola A, Negi D, Manchanda M, Kesharwani RK (2022) Chapter 27: Bioinformatics and biological data mining. Academic Press, New York, pp 457–471. https://doi.org/10.1016/B978-0-323-89775-4.00019-5

Hardcastle TJ, Kelly KA (2010) baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinform 11:422

Harmanci A, Rozowsky J, Gerstein M (2014) MUSIC: identification of enriched regions in ChIP-Seq experiments using a mappability-corrected multiscale signal processing framework. Genome Biol 15:474

Hashimoto TB, Edwards MD, Gifford DK (2014) Universal count correction for high-throughput sequencing. PLOS Comput Biol 10(3):e1003494. https://doi.org/10.1371/journal.pcbi.1003494

Heinz S et al (2010) Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. Mol Cell 38:576–589

Hernandez D, Francois P, Farinelli L, Osteras M, Schrenzel J (2008) De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. Genome Res 18: 802–809

Hertel J, Stadler PF (2006) Hairpins in a Haystack: recognizing microRNA precursors in comparative genomics data. Bioinformatics 22(14):e197–e202. https://doi.org/10.1093/bioinformatics/btl257

Hoffmann S, Otto C, Doose G, Tanzer A, Langenberger D, Christ S et al (2014) A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. Genome Biol 15(2):R34

Homer N, Merriman B, Nelson SF (2009) BFAST: an alignment tool for large scale genome resequencing. PLoS One 4(11):e7767

Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K (2007) WoLF PSORT: protein localization predictor. Nucleic Acids Res 35:W585–W587. https://doi.org/10.1093/nar/gkm259

Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5:e1000529

Hrdlickova R, Toloue M, Tian B (2017) RNA-Seq methods for transcriptome analysis. Wiley Interdiscip Rev RNA 8(1):1364. https://doi.org/10.1002/wrna.1364

Hu R, Sun X (2016) lncRNATargets: a platform for lncRNA target prediction based on nucleic acid thermodynamics. J Bioinform Comput Biol 14(4):1650016. https://doi.org/10.1142/S0219720016500165

Huang Y-A, Huang Z-A et al (2019) Predicting lncRNA–miRNA interaction via graph convolution auto-encoder. Front Genet 10:758

Hubley R et al (2016) The Dfam database of repetitive DNA families. Nucleic Acids Res 44:D81–D89

Izuogu OG, Alhasan AA, Alafghani HM, Santibanez-Koref M, Elliott DJ, Jackson MS (2016) PTESFinder: a computational method to identify post-transcriptional exon shuffling (PTES) events. BMC Bioinform 17:31

Jha A, Shankar R (2013) miReader: discovering novel miRNAs in species without sequenced genome. PLoS One 8(6):e66857. https://doi.org/10.1371/journal.pone.0066857

Jühling F, Kretzmer H, Bernhart SH, Otto C, Stadler PF, Hoffmann S (2016) Metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. Genome Res 26(2):256–262. https://doi.org/10.1101/gr.196394.115

Kajitani R, Toshimoto K, Noguchi H et al (2014) Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res 24(8):1384–1395. https://doi.org/10.1101/gr.170720.113

Kajitani R, Yoshimura D, Okuno M et al (2019) Platanus-allee is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous regions. Nat Commun 10:1702. https://doi.org/10.1038/s41467-019-09575-2

Kalvari I et al (2020) Rfam 14: expanded coverage of metagenomic, viral and microRNA families. Nucleic Acids Res 49(D1):D192–D200

Kang YJ et al (2017) CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Res 45(W1):W12–W16. https://doi.org/10.1093/nar/gkx428

Karagkouni D, Paraskevopoulou MD et al (2019) DIANA-LncBase v3: indexing experimentally supported miRNA targets on non-coding transcripts. Nucleic Acids Res 48:D101–D110

Katz Y, Wang ET, Airoldi EM, Burge CB (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat Methods 7:1009–1015

Kearse M, Moir R, Wilson A et al (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12): 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kent WJ (2002) BLAT—the BLAST-like alignment tool. Genome Res 12(4):656–664

Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E (2007) The role of site accessibility in microRNA target recognition. Nat Genet 39:1278–1284. https://doi.org/10.1038/ng2135

Khalil AM, Rinn JL (2011) RNA–protein interactions in human health and disease. Semin Cell Dev Biol 22:359–365

Khan A, Singh K et al (2022) Whole-genome-based web genomic resource for water buffalo (*Bubalus bubalis*). Front Genet 13:9741. https://doi.org/10.3389/fgene.2022.809741

Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12(4):357–360. https://doi.org/10.1038/nmeth.3317

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22(3):568–576

Kofler R, Schlotterer C, Lelley T (2007) SciRoKo: a new tool for whole genome microsatellite search and investigation. Bioinformatics 23:1683–1685. https://doi.org/10.1093/bioinformatics/btm157

Kong L et al (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res 35:W345–W349. https://doi.org/10.1093/nar/gkm391

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27(5):722–736. https://doi.org/10.1101/gr.215087.116

Korf I (2004) Gene finding in novel genomes. BMC Bioinform 5:59

Kozomara A, Griffiths-Jones S (2010) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 39(Database):D152–D157. https://doi.org/10.1093/nar/gkq1027

Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M et al (2005) Combinatorial microRNA target predictions. Nat Genet 37: 495–500. https://doi.org/10.1038/ng1536

Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305(3): 567–580. https://doi.org/10.1006/jmbi.2000.4315

Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27(11):1571–1572. https://doi.org/10.1093/bioinformatics/btr167

Kurdyukov S, Bullock M (2016) DNA methylation analysis: choosing the right method. Biology 5(1):E3. https://doi.org/10.3390/biology5010003

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25

Le SQ, Durbin R (2010) SNP detection and genotyping from low-coverage sequencing data on multiple diploid samples. Genome Res 21(6):952–960. https://doi.org/10.1101/gr.113084.110

Lee B, Baek J, Park S, Yoon S (2016) deepTarget: End-to-end learning framework for MicroRNA target prediction using deep recurrent neural networks. In: Proceedings of the 7th ACM international conference on bioinformatics, computational biology, and health informatics, Seattle, WA, USA, 2 October–5 October 2016; pp 434–442

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England) 25:1754–1760

Li H, Ruan J, Durbin RM (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res 18:1851–1858

Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Yang H, Wang J (2009a) De novo assembly of human genomes with massively parallel short read sequencing. Genome Res 20:265–272

Li H et al (2009b) The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25: 2078–2079

Li R et al (2009c) SNP detection for massively parallel whole-genome resequencing. Genome Res 19:1124–1132

Li R et al (2009d) SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25: 1966–1967

Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J (2009e) SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25:1966–1967

Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 34:816–834. https://doi.org/10.1002/gepi.20533

Li Z et al (2012) Comparison of the two major classes of assembly algorithms: overlap–layout–consensus and de-bruijn-graph. Brief Funct Genom 11(1):25–37. https://doi.org/10.1093/bfgp/elr035

Li J, Ma W, Zeng P, Wang J, Geng B, Yang J, Cui Q (2015) LncTar: a tool for predicting the RNA targets of long noncoding RNAs. Brief Bioinform 16(5):806–812. https://doi.org/10.1093/bib/bbu048

Li J, Yang Z, Wang D, Li Z (2022) WAFNRLTG: a novel model for predicting LncRNA target genes based on weighted average fusion network representation learning method. Front Cell Dev Biol 9:820342. https://doi.org/10.3389/fcell.2021.820342

Liao Y, Smyth GK, Shi W (2013) The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res 41(10):e108. https://doi.org/10.1093/nar/gkt214

Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30(7):923–930. https://doi.org/10.1093/bioinformatics/btt656

Licatalosi DD, Mele A, Fak JJ, Ule J, Kayikci M, Chi SW, Clark TA, Schweitzer AC, Blume JE, Wang X et al (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature 456:464–469

Lin MF, Jungreis I, Kellis M (2011) PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. Bioinformatics 27(13):i275–i282. https://doi.org/10.1093/bioinformatics/btr209

Lisch D (2013) How important are transposons for plant evolution? Nat Rev Genet 14(1):49–61. https://doi.org/10.1038/nrg3374

Liu C-M, Wong T, Wu E, Luo R, Yiu S-M, Li Y, Wang B, Yu C, Chu X, Zhao K, Li R, Lam T-W (2012) SOAP3: ultra-fast GPU-based parallel alignment tool for short reads. Bioinformatics 28(6):878–879

Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B (2012) RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res 40:W622–W627

Lunter G, Goodson M (2010) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res 21(6):936–939. https://doi.org/10.1101/gr.111120.110

Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ (2009) GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinform 10:161

Luo R, Liu B et al (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 1(1):18. https://doi.org/10.1186/2047-217X-1-18

Luo R, Wong T, Zhu J, Liu C-M, Zhu X, Wu E, Lee L-K, Lin H, Zhu W, Cheung DW, Ting H-F, Yiu S-M, Peng S, Yu C, Li Y, Li R, Lam TW (2013) SOAP3-dp: fast, accurate and sensitive GPU-based short read aligner. PLoS One 8(5):e65632

Maccallum I, Przybylski D, Gnerre S, Burton J, Shlyakhter I, Gnirke A, Malek J, McKernan K, Ranade S, Shea TP, Williams L, Young S, Nusbaum C, Jaffe DB (2009) ALLPATHS 2: small genomes assembled accurately and with high continuity from short paired reads. Genome Biol 10:R103

Maia LC, Palmieri DA, Souza VQ, Kopp MM, Carvalho FIF, Oliveira AC (2008) SSR locator: tool for simple sequence repeat discovery integrated with primer design and PCR simulation. Int J Plant Genom 2008:412696

Majoros WH, Pertea M, Salzberg SL (2004) TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics 20:2878–2879

Manchanda N, Portwood JL, Woodhouse MR et al (2020) GenomeQC: a quality assessment tool for genome assemblies and gene structure annotations. BMC Genomics 21:193. https://doi.org/10.1186/s12864-020-6568-2

Manni M, Berkeley MR, Seppey M, Zdobnov EM (2021) BUSCO: assessing genomic data quality and beyond. Curr Protoc 1:e323

Mapleson D, Accinelli GG, Kettleborough G, Wright J, Clavijo BJ (2017) KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics 33(4):574–576. https://doi.org/10.1093/bioinformatics/btw663

Margulies M, Egholm M et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380

Mas-Ponte D, Carlevaro-Fita J, Palumbo E, Pulido TH, Guigo R, Johnson R (2017) LncATLAS database for subcellular localization of long noncoding RNAs. RNA 23:1080–1087. https://doi.org/10.1261/rna.060814.117

McKenna A, Hanna M, Banks E, Sivachenko A et al (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303

McLean CY et al (2010) GREAT improves functional interpretation of *cis*-regulatory regions. Nat Biotechnol 28:495–501

Medema MH, Blin K, Cimermancic P et al (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res 39(Web Server issue):W339–W346. https://doi.org/10.1093/nar/gkr466

Memczak S, Jens M, Elefsinioti A et al (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 495:333–338

Merino GA, Conesa A, Fernández EA (2019) A benchmarking of workflows for detecting differential splicing and differential expression at isoform level in human RNA-seq studies. Brief Bioinform 20(2):471–481. https://doi.org/10.1093/bib/bbx122

MINE SSR (2005). http://www.genome.clemson.edu/resources/online_tools/ssr

MIRA (2023). chevreux.org/projects_mira.html

Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, Lim B, Rigoutsos I (2006) A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. Cell 126:1203–1217. https://doi.org/10.1016/j.cell.2006.07.031

Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A (2021) Pfam: the protein families database in 2021. Nucleic Acids Res 49(D1):D412–D419. https://doi.org/10.1093/nar/gkaa913

Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, Guigo R, Dermitzakis ET (2010) Transcriptome genetics using second generation sequencing in a Caucasian population. Nature 464:773–777

Muñoz-Espinoza C, Di Genova A, Sánchez A et al (2020) Identification of SNPs and InDels associated with berry size in table grapes integrating genetic and transcriptomic approaches. BMC Plant Biol 20:365. https://doi.org/10.1186/s12870-020-02564-4

Myers EW, Sutton GG et al (2000) A whole-genome assembly of Drosophila. Science 287:2196–2204

Nagarajan N, Pop M (2013) Sequence assembly demystified. Nat Rev Genet 14(3):157–167

Nam J-W, Rissland OS, Koppstein D, Abreu-Goodger C, Jan CH, Agarwal V, Yildirim MA, Rodriguez A, Bartel DP (2014) Global analyses of the effect of different cellular contexts on microRNA targeting. Mol Cell 53:1031–1043. https://doi.org/10.1016/j.molcel.2014.02.013

Nawrocki EP, Eddy SR (2013) Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29:2933–2935

Neary JL, Carless MA (2020) Methylated DNA immunoprecipitation sequencing (MeDIP-seq): principles and applications. Epigenet Methods 18:157–179. https://doi.org/10.1016/b978-0-12-819414-0.00009-4

Nicol JW, Helt GA, Blanchard SG et al (2009) The integrated genome browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics 25(20):2730–2731. https://doi.org/10.1093/bioinformatics/btp472

Nielsen R, Korneliussen T, Albrechtsen A et al (2012) SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. PLoS One 7(7):e37558. https://doi.org/10.1371/journal.pone.0037558

Pandey RV, Pabinger S, Kriegner A et al (2016) ClinQC: a tool for quality control and cleaning of Sanger and NGS data in clinical research. BMC Bioinform 17:56. https://doi.org/10.1186/s12859-016-0915-y

Park Y, Wu H (2016) Differential methylation analysis for BS-seq data under general experimental design. Bioinformatics 32(10):1446–1453. https://doi.org/10.1093/bioinformatics/btw026

Park S, Min S, Choi H, Yoon S (2016) deepMiRGene: deep neural network based precursor MicroRNA prediction. https://arxiv.org/abs/1605.00017

Patro R, Mount SM, Kingsford C (2014) Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. Nat Biotechnol 32(5):462–464

Peng Y et al (2010) IDBA-A practical iterative de Bruijn graph de novo assembler. RECOMB, Lisbon

Peng C, Han S, Zhang H, Li Y (2019) RPITER: a hierarchical deep learning framework for ncRNA–protein interaction pre-diction. Int J Mol Sci 20:1070

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 33(3):290–295. https://doi.org/10.1038/nbt.3122

Pertea M, Shumate A, Pertea G, Varabyou A, Breitwieser FP, Chang YC et al (2018) CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. Genome Biol 19:208. https://doi.org/10.1186/s13059-018-1590-2

Pevzner PA, Tang H, Tesler G (2004) De novo repeat classification and fragment assembly. Genome Res 14:1786–1796

Pimentel H, Bray NL, Puente S, Melsted P, Pachter L (2017) Differential analysis of RNA-seq incorporating quantification uncertainty. Nat Methods 14(7):687–690. https://doi.org/10.1038/nmeth.4324

Price AL, Jones NC, Pevzner PA (2005) De novo identification of repeat families in large genomes. Bioinformatics 21(suppl. 1):i351–i358

Quevillon E, Silventoinen V, Pillai S et al (2005) InterProScan: protein domains identifier. Nucleic Acids Res 33(Web Server issue):W116–W120. https://doi.org/10.1093/nar/gki442

Rashid NU, Giresi PG, Ibrahim JG, Sun W, Lieb JD (2011) ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions. Genome Biol 12:R67

Reczko M, Maragkakis M, Alexiou P, Grosse I, Hatzigeorgiou AG (2012) Functional microRNA targets in protein coding sequences. Bioinformatics 28:771–776. https://doi.org/10.1093/bioinformatics/bts043

RepeatMasker (2019). https://www.repeatmasker.org/cgi-bin/WEBRepeatMasker

REPET (2019). https://urgi.versailles.inra.fr/Tools/REPET

Riley KJ, Steitz JA (2013) The "observer effect" in genome-wide surveys of protein–RNA interactions. Mol Cell 49:601–604

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140

Salzberg SL (2019) Next-generation genome annotation: we still struggle to get it right. Genome Biol 20(1):92. https://doi.org/10.1186/s13059-019-1715-2

Schmid MW, Grossniklaus U (2015) Rcount: simple and flexible RNA-Seq read counting. Bioinformatics 31(3):436–437. https://doi.org/10.1093/bioinformatics/btu680

Schweikert G, Behr J, Zien A, Zeller G, Ong CS et al (2009) mGene.web: a web service for accurate computational gene finding. Nucleic Acids Res 37:W312–W316

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11): 2498–2504

Shen Y, Wan Z, Coarfa C, Drabek R, Chen L, Ostrowski EA, Liu Y, Weinstock GM, Wheeler DA, Gibbs RA, Yu F (2010) A SNP discovery method to assess variant allele probability from next generation resequencing data. Genome Res 20(2):273–280

Shu L, Arneson D, Yang X (2018) Bioinformatics principles for deciphering cardiovascular diseases. In: Encyclopedia of cardiovascular research and medicine. Elsevier, Amsterdam, pp 273–292. https://doi.org/10.1016/B978-0-12-809657-4.99576-0

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I (2009) ABySS: a parallel assembler for short read sequence data. Genome Res 19:1117–1123

Slotkin RK, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet 8(4):272–285. https://doi.org/10.1038/nrg2072

Smith AD, Xuan Z, Zhang MQ (2008) Using quality scores and longer reads improves accuracy of Solexa read mapping. BMC Bioinform 9:128

Sommer DD, Delcher AL, Salzberg SL et al (2007) Minimus: a fast, lightweight genome assembler. BMC Bioinform 8:64. https://doi.org/10.1186/1471-2105-8-64

Song X, Zhang N, Han P, Moon BS, Lai RK, Wang K et al (2016) Circular RNA profile in gliomas revealed by identification tool UROBORUS. Nucleic Acids Res 44(9):e87

Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res 33(Web Server issue):W465–W467. https://doi.org/10.1093/nar/gki458

Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM (2005) Animal microRNAs confer robustness to gene expression and have a significant impact on 3′UTR evolution. Cell 123: 1133–1146

Stockwell PA et al (2014) DMAP: differential methylation analysis package for RRBS and WGBS data. Bioinformatics 30(13):1814–1822. https://doi.org/10.1093/bioinformatics/btu126

Sun L et al (2013) Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Res 41(17):e166. https://doi.org/10.1093/nar/gkt646

Sutton GG, White O, Adams MD, Kerlavage AR (1995) TIGR assembler: a new tool for assembling large shotgun sequencing projects. Genome Sci Technol 1:9–19

Szabo L, Morey R, Palpant NJ, Wang PL, Afari N, Jiang C et al (2015) Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development. Genome Biol 16:126

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11(8):1441–1452. https://doi.org/10.1101/gr.184001

Tempel S, Tahi F (2012) A fast ab-initio method for predicting miRNA precursors in genomes. Nucleic Acids Res 40(11):e80. https://doi.org/10.1093/nar/gks146

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development of cDNA derived microsatellite markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422

Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14(2):178–192

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111

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 28:511–515

Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol 31(1):46–53. https://doi.org/10.1038/nbt.2450

USADELLAB (2016). http://www.usadellab.org/cms/?page=trimmomatic

Vieira ML, Santini L, Diniz AL, de Munhoz CF (2016) Microsatellite markers: what they mean and why they are so useful. Genet Mol Biol 39(3):312–328. https://doi.org/10.1590/1678-4685-GMB-2016-0027

Wang X, Cairns MJ (2014) SeqGSEA: a bioconductor package for gene set enrichment analysis of RNA-Seq data integrating differential expression and splicing. Bioinformatics 30(12): 1777–1779

Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, He X, Mieczkowski P, Grimm SA, Perou CM et al (2010) MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res 38:e178

Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, He X, Miecz-kowski P, Grimm SA, Perou CM et al (2010a) MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res 38:e178

Wang L, Feng Z, Wang X, Wang X, Zhang X (2010b) DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics 26:136–138

Wei Z, Wang W, Hu P, Lyon GJ, Hakonarson H (2011) SNVer: a statistical tool for variant calling in analysis of pooled or individual next-generation sequencing data. Nucleic Acids Res 39(19): e132

Westholm JO, Miura P, Olson S, Shenker S, Joseph B, Sanfilippo P, Celniker SE, Graveley BR, Lai EC (2014) Genome-wide analysis of Drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. Cell Rep 9:1966–1980

Wicker T, Sabot F, Hua-Van A et al (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982. https://doi.org/10.1038/nrg2165

Wilson GA, Beck S (2016) Computational analysis and integration of MeDIP-seq methylome data. In: Kulski JK (ed) Next generation sequencing: advances, applications and challenges. InTechOpen, Rijeka

Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26:873–881

Wu TD, Watanabe CK (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21(9):1859–1875. https://doi.org/10.1093/bioinformatics/bti310

Xing H, Mo Y, Liao W et al (2012) Genome-wide localization of protein-DNA binding and histone modification by a Bayesian change-point method with Chip-Seq data. PLoS Comput Biol 8: e1002613

Xiong Q, Mukherjee S, Furey TS (2014) GSAASeqSP: a toolset for gene set association analysis of RNA-Seq data. Sci Rep 4:6347

Xu H, Wei CL, Lin F, Sung WK (2008) An HMM approach to genome-wide identification of differential histone modification sites from ChIP-seq data. Bioinformatics 24(20):2344–2349. https://doi.org/10.1093/bioinformatics/btn402

Yang G (2013) Mite digger, an efficient and accurate algorithm for genome wide discovery of miniature inverted repeat transposable elements. BMC Bioinform 14(1):186

Yang X, Li L (2011) miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. Bioinformatics 27(18):2614–2615. https://doi.org/10.1093/bioinformatics/btr430

Yang X, Liu D, Liu F et al (2013) HTQC: a fast quality control toolkit for Illumina sequencing data. BMC Bioinform 14:33. https://doi.org/10.1186/1471-2105-14-33

Yang LA, Chang YJ, Chen SH et al (2019) SQUAT: a sequencing quality assessment tool for data quality assessments of genome assemblies. BMC Genomics 19:238. https://doi.org/10.1186/s12864-019-5445-3

Yao Z, You FM et al (2020) Evaluation of variant calling tools for large plant genome re-sequencing. BMC Bioinform 21:360. https://doi.org/10.1186/s12859-020-03704-1

Ye C, Ji G, Liang C (2016) Detectmite: a novel approach to detect miniature inverted repeat transposable elements in genomes. Sci Rep 6:19688

Yi H-C, You Z et al (2018) A deep learning framework for robust and accurate prediction of ncRNA–protein interactions using evolutionary information. Mol Ther Nucleic Acids 11:337–344

Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol 11(2):R14

Yuan C, Sun Y (2013) RNA-CODE: a noncoding RNA classification tool for short reads in NGS data lacking reference genomes. PLoS One 8(10):e77596. https://doi.org/10.1371/journal.pone.0077596

Zang C, Schones DE, Zeng C et al (2009) A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics 25:1952–1958

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18(5):821–829. https://doi.org/10.1101/gr.074492.107

Zerbino DR, McEwen GK, Margulies EH, Birney E (2009) Pebble and rock band: heuristic resolution of repeats and scaffolding in the velvet short-read de novo assembler. PLoS One 4:e8407

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS et al (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9:R137

Zhang T, Luo Y, Liu K, Pan L, Zhang B, Yu J, Hu S (2011) BIGpre: a quality assessment package for next-generation sequencing data. Genom Proteom Bioinform 9(6):238–244. https://doi.org/10.1016/S1672-0229(11)60027-2

Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L (2014) Complementary sequence-mediated exon circularization. Cell 159:134–147

Zhang H, He L, Cai L (2018) Transcriptome sequencing: RNA-Seq. Methods Mol Biol 1754:15–27. https://doi.org/10.1007/978-1-4939-7717-8_2

Zhu LJ et al (2010) ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinform 11:237

Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA (2013) The MaSuRCA genome assembler. Bioinformatics 29(21):2669–2677. https://doi.org/10.1093/bioinformatics/btt476

Chapter 20
Application of AI/ML Approaches for Livestock Improvement and Management

Manisha Malhotra, Akanksha Jaiswar, Abhimati Shukla, Nivedita Rai, Ashish Bedi, Mir Asif Iquebal, Sarika Jaiswal, Dinesh Kumar, and Anil Rai

Abstract At present, livestock is one of the demanding agricultural subsectors in developing countries due to the growing population, advanced urbanization, and escalating demand for animal food. In addition, livestock makes a vital contribution to the economy of rural and urban areas by providing value-added goods. However, the limitations of natural resources and land make efficient growth of animals per hectare and their daily management a higher challenge. Therefore, to overcome these challenging problems and to find the optimal solutions to maximize the efficient livestock production by minimizing the cost and managing the health and welfare of livestock, new emerging technologies such as (AI) and machine learning (ML) can be employed as they have potentially been utilized in transforming various other sectors. Therefore, the current chapter explores the role of AI and ML in improving and managing the livestock by reviewing their applications in lowering the livestock production cost, enhancing their production per hectare, improving the health of animals, and resource management.

Manisha Malhotra and Akanksha Jaiswar contributed equally to this work.

M. Malhotra
CSIR-Indian Institute of Petroleum, Dehradun, India

A. Jaiswar · M. A. Iquebal · S. Jaiswal (✉) · D. Kumar · A. Rai
Division of Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India
e-mail: ma.iquebal@icar.gov.in; sarika@icar.gov.in; dinesh.kumar@icar.gov.in; anil.rai@icar.gov.in

A. Shukla
Harcourt Butler Technical University, Kanpur, India

N. Rai
School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India

A. Bedi
School of Life and Environmental Sciences, Deakin University, Geelong, Victoria, Australia

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_20

Keywords Machine learning · Algorithm · Population · Economy

20.1 Introduction

As the global population is increasing, there is advancement in urbanization, and enhanced demand for animal food and products, expanding the need to produce more animals per hectare and also the advancement in food processing industrialization (Delgado 2005). Moreover, livestock production not only helps increase the economy by yielding value-added products, but can also be utilized for transportation and fuel. Therefore, livestock production is becoming one of the rapidly progressing agricultural subsectors in developing countries and help increase their GDP. However, the complexities in livestock farming, due to reduced land and natural resources, and highly compromised traditional livestock management, make it really challenging for farmers and stakeholders to fulfil the upward demand for livestock-based products and food.

Moreover, these traditional methods are time consuming and require a huge amount of labor and laborers. Hence, to overcome these limitations, traditional methods of farming need to be replaced by more advanced technologies aiming to increase the growth of more animals per hectare, improve livestock management, reduce environmental impact by proper natural resource management, and improve the health and feed of livestock (Pomar et al. 2011; Garcia et al. 2020). More recently, technologies based on big data such as (AI), machine learning (ML), and sensors are overpowering other traditional technologies, and their involvement in livestock production and management is no different. Since these big data-mediated technologies function on bulk collection of varying datasets, the implementation of these mechanistic models in livestock farming requires the collection of diverse datasets such as voice signals, visuals, movement, and behavior of farm animals, data of local weather and air quality (Abraham and Pingali 2020).

Furthermore, the algorithms based on these technologies, correlate, segregate, and cluster the patterns within the datasets depending on the data provided and help farmers monitor the farm animals, also help them detect various abnormalities and diseases in livestock at early stages. Figure 20.1 illustrates how the big data is collected, processed, and further clustered together using AI/ML technology.

Therefore, utilizing these advanced technologies may efficiently enhance our knowledge in extracting valuable information from data and elevate our understanding towards complex animal systems (Ellis et al. 2020).

Furthermore, the current scenario helps farmers and stakeholders to utilize internet of things, cloud computing, ML, and AI in controlling farm system. Various devices and machines with built-in intelligence and well equipped with sensors plays potential role in management and analysis of farm animals, therefore help farmers analyze and predict any kind of abnormalities in the movement and behavior of livestock, resulting in disease detection and its prevention even during pre-clinical

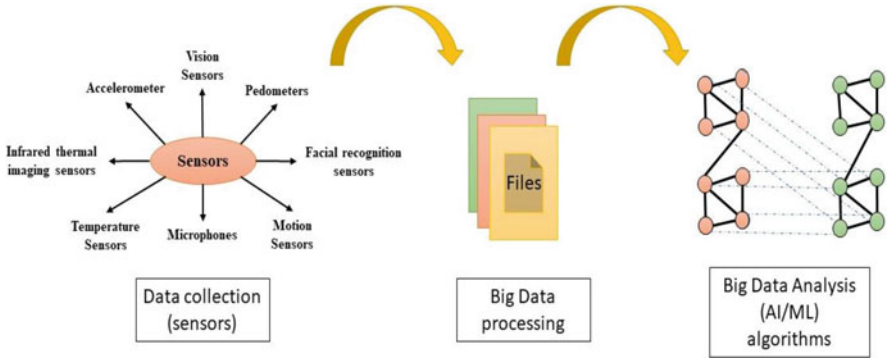

Fig. 20.1 Illustration of big data collection through sensors, its processing and segregation using AI/ML technologies

stages and enable a smaller number of farmers to monitor large number of livestock quickly which results in reducing their cost of production. For instance, a number of sensors with ML/AI technologies have been generated to identify and track the livestock and to analyze them based on their behavioral patterns in real time. Not only this, but also farmers can track the sleep cycles of animals and the quality of air in the livestock shelters; for example, air sensors can help prognosticate the commencement of coccidiosis in poultry farms and help farmers prevent the outbreak of the disease, as these sensors detect the changes in concentration of volatile organic compounds, resulted due to the infection in birds by continuously analyzing the quality of air (Borgonovo et al. 2020).

Hence, these technologies not only prevent the lives of livestock but also impede financial losses. Furthermore, AI/ML algorithms can analyze the data, obtained from automated measurement of different physiological and environmental variables; for instance, internal temperature, surface temperature, respiration rate, sweating rate, walking pattern, behavior and weight of animals which may further help earlier detection of injury, discomfort, heat stress, metabolic distress, and illness in livestock (Zhang et al. 2020). For example, these advanced data-mediated technologies can help earlier diagnosis and prevention of various contagious diseases such as swine flu and other diseases such as mastitis which can be detected by various traditional methods; for instance, mastitis, an udder disease in cows, results in low quantity and quality of milk production that can be diagnosed manually by somatic cell counts (SSC) and electric conductivity (EC) readings which are mostly unreliable and RT-PCR techniques which on the other hand are reliable but are costly and cannot be implicated on large scale (Gertz et al. 2020; Hidalgo et al. 2018).

Therefore, sensors and algorithms based on AI/ML technologies have proved to be potentially beneficial in getting quick, reliable, and more accurate data, which help earlier diagnosis and prevention of these contagious diseases in cost-effective manner. In yet another study, researchers showed how AI and ML technologies can be utilized to detect the infection of *Clostridium perfringens* in birds by providing

the data of recorded broiler vocalizations from both the healthy and diseased birds and then the differentiation was analyzed utilizing an artificial neural network (ANN) model by identifying and correlating five data clusters (Sadeghi et al. 2015). Furthermore, in various studies, researchers utilize various models and algorithms to monitor the body weight of different farm animals so as to predict an appropriate feed allocation to these animals. For example, body weight of the growing pigs was monitored using machine vision depending on the analysis of visual images so as to allocate them proper feed. Besides determining feed efficiency, certain factors need to be monitored such as quantity of feed consumed, gain in weight of livestock, and the quantity of milk and eggs yielded. Hence, there are various sensors, for instance, RGB—D cameras that can help determine the quantity of feed to be consumed by individual cows and algorithms utilized such as TDIDT, ENET, SSD, ARIMA, and CNNs, which can help farmers determine the consumption of feed and also can optimize the cost of feed (Piles et al. 2019; Nikoloski et al. 2019; da Rosa Righi et al. 2020). ML/AI technology can further determine the performance of farm animals by quantitating the yield of milk and body condition score (BCS), and also help estimating the breeding time, calving time, and the reproductive performance of livestock.

Moreover, identification and recognition of an individual animal in a herd has now become easier with the help of various facial recognition methods; for instance, in a study, three methods such as VGG-face model (Parkhi et al. 2015), Fisher faces (Belhumeur et al. 1997), and convolutional neural networks were used simultaneously which help recognize the faces of individual pigs with an efficiency of 96.7%. In yet another example, PANSNet-5 recognition model, when combined with face recognition system, the efficiency of detecting cow's face increased by 98.3% (Liu et al. 2018). Hence, application of different models and algorithms based on these technologies can help maintain appropriate number of livestock per farm, improve health and disease resistance in livestock, help boost their feed efficiency and nutritional status, facilitate the management of resources, and help conduct dairy energy analysis, which otherwise become cumbersome for farmers alone to take care of. Also, various models and algorithms have been taken into account to manage pasture, nutrition, and feed factors of livestock, for analyzing the effect of feed on pH of rumen of cattle, their milk production, digestion, and also gas emission (Bannink et al. 2011).

Therefore, this chapter explores the role of various algorithms and methodologies based on AI and ML in improving the management and production of livestock and help farmers reduce production cost of these livestock.

20.2 Limitations of Traditional Livestock Farming Methods

Animals are raised for food and other purposes, including the production of leather, wool, and even fertilizer. Although cattle or dairy cows, chickens, goats, pigs, horses, and sheep are the main livestock that fall under this category of farming, other animals such as donkeys, mules, rabbits, and insects such as bees are also becoming more and more relevant (National Research Council 2003). Ever since people began domesticating animals to improve their quality of life, livestock farming has been a significant aspect of human society for the majority of recorded history. But as with most forms of farming, such as agriculture, livestock farming too has intensified, particularly in recent decades (Thornton 2010).

This has allowed livestock farming products more broadly accessible and less expensive to purchase, which is especially significant if you consider that essentials like milk, honey, eggs, and meat are all products of livestock farming. Livestock farming is frequently referred to as "factory farming," however, significant concerns have been raised over food safety, animal welfare, and environmental implications due to the techniques of intensive livestock farming on numerous occasions (Kunkel 2003).

We cannot rule out the increasing animal welfare, environmental, and health concerns while talking about traditional livestock farming. Traditional livestock farming is cost effective but the reduced cost of these techniques often has an impact on the health and well-being of the animals. Moreover, the areas in which the livestock animals are kept are very much crowded and have poor hygiene and living conditions which results in the animals being more susceptible to diseases. In developing countries, some livestock-related diseases cause illness to about 2.4 billion human population.

Farmers, in order to control the livestock-related diseases, use extensive amount of antibiotics and long-term use of antibiotics lead to the development of antibiotic resistance in bacteria and pathogens. Lack of awareness among the farmers of the importance of growing the better-quality of forage species, and low practice of cultivated forage due to the shortage of land, leads to the shortage of feed and fodders for the animals which results in poor livestock health. Also, the lack of knowledge of balanced feed also contributes to the poor health of the animals (Kumar et al. 2012). Seasonal feed shortages particularly in the dry season were pointed out as the major constraints to increasing ruminant productivity in developing countries (Ayele et al. 2021). Other shortcomings of traditional livestock farming are the unavailability of water in the dry season and low productive capacity of local livestock breeds (Umutoni et al. 2015). As a consequence, livestock is generally malnourished during this time and thus becomes more susceptible to diseases. Yet, another disadvantage of traditional livestock farming is the poor reproductive performance and low productivity (low milk production) of the locally adapted breeds as compared to elite varieties (Fig. 20.2). Some other problems reported by the farmers include the conflict between herders and farmers. These constraints are

Fig. 20.2 Various concerns related to traditional livestock breeding

symptomatic of the extensive livestock production system which is still dominant in developing countries (Kebreab et al. 2005).

Recently, some AI-based scientific studies have been done to address issues with animal welfare and human health (Ezanno et al. 2021). However, no comprehensive study was carried out to evaluate the potential of AI in addressing the many application issues in a number of animal species, including ruminants, pigs, and poultry. Interestingly, this chapter will offer a thorough evaluation of scientific research advances in AI-related animal farming based on cumulative studies based on previous studies in order to close the knowledge gap.

AI technology plays a significant role in helping smart farming in the area of improving animal health and welfare in order to realize good economic gain in intensive animal husbandry (Mohamed et al. 2021). The following are some important applications of ML and AI in livestock farming:

20.3 Application of ML and AI

20.3.1 Animal Farming

The application of AI in animal farming is becoming a data-centric business these days. Through their sensors and AI technology, latest dairy is applying intelligence to raise animals for meat, milk, fiber, eggs, and other products. They provide a variety of sensors, such as those for calving and heat detection as well as health monitoring devices like the Sense Time Solution sensor, which tracks a cow's daily activities like rumination, eating, and walking patterns. Numerous sensors are now available to enable farmers monitor changes in animal movements, food intake, sleep cycles, and even air quality in animal habitats, shelters. When combined with such intelligence tools like sensors, the prior information offered by them leads to proactive responses to solve the issues. Additionally, to the ability to record facts on health, reproduction, and the sensor also offers farmers nutrition with answers

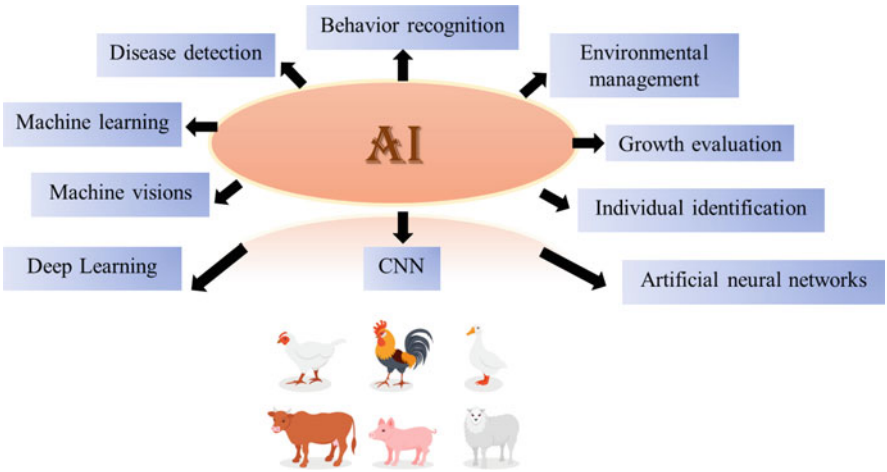

Fig. 20.3 Application of AI and ML in livestock's farming

specific to each cow (Linaza et al. 2021). Also, the AI wearable devices are being used to collect real-time data like health of the animals, vaccination details, and insemination information (Fig. 20.3).

The various uses of ML and AI in animal farming are described as follows:

- **Automated milking using AI:** In milking, AI can be used for analyzing the quality of milk and detect the abnormalities if any in the product.
- **AI in improvement of feed quality:** When compared to traditional manual harvesting methods, the usage of robotics and AI is highly efficient and reduces harvesting time. Additionally, the automated equipment continuously determines the yield and moisture content of the harvest of cereals.
- **AI in food supply chain:** Blockchain can link every link in the supply chain, from the producer to the consumer, enabling food safety and traceability. From the perspective of agriculture and food, presenting this kind of information to consumers will become a competitive advantage and might not be as difficult in dairy as it is in other agricultural fields, like beef, which changes ownership more frequently (Dora et al. 2021) (Fig. 20.4).
- **Improving animal health using facial recognition systems:** There are numerous beneficial applications, including assisting us in understanding the animal's emotional and attentional states. For instance, scientists can now reasonably predict an animal's mood and amount of enthusiasm by observing its ear and eye movements. It might assist us in controlling animal pain symptoms. Further investigation may turn up wounds, illnesses, or even signs of predator assaults.
- **Robotic system to deliver vaccines:** For dairy farms to have a viable economic future and to reach 100% utilization of a compliance rate, modern dairy farms robotic injection device for vaccine delivery and contraceptive drugs for domesticating the livestock on the dairy farm. The robotic apparatus has a dairy

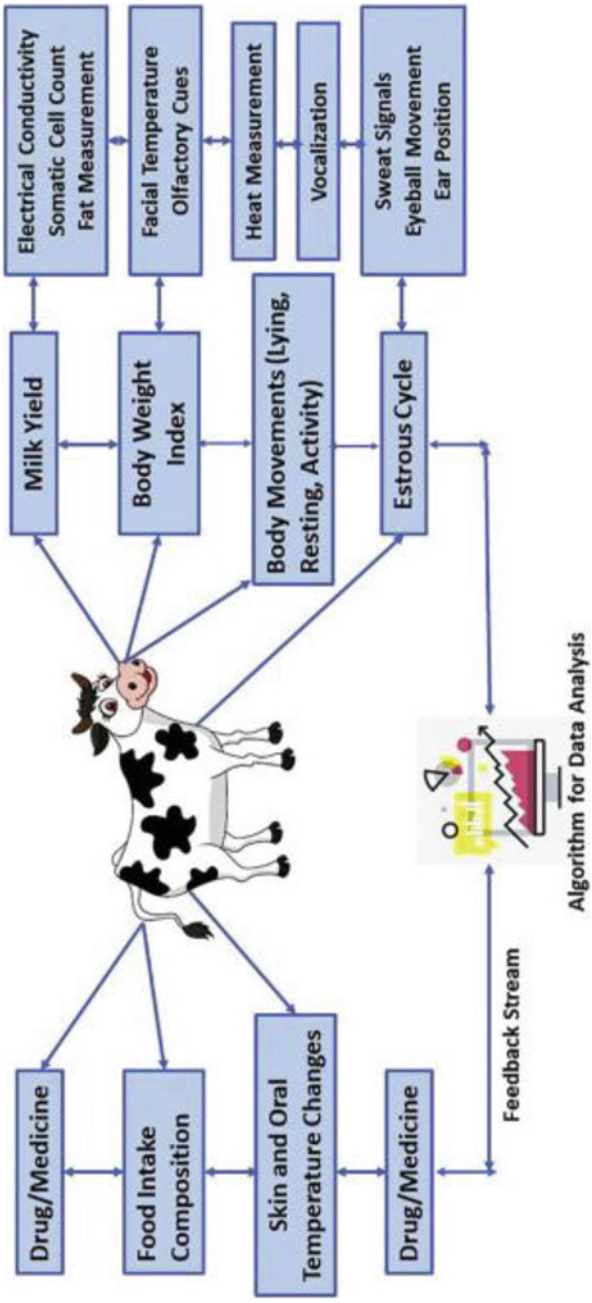

Fig. 20.4 This schema shows a possible interpretation of data using ML algorithms to produce the best growing circumstances for dairy farming. Source figure taken from Neethirajan 2020

automation built-in system in use today. Automated injection device reads the embedded RFID tags on the receives health-related information from a cow's ear and the cow's immunization history (Haldar et al. 2022). Should a cow need to be injected, so that direction is taken location and the position of the injection mechanism to inject the drug into the cow's neck.

- **AI for detection of oestrus:** The motion sensor-equipped collar that is fastened to the cow's neck continuously gathers all kinds of information about the animal. The dairy automation system's AI components analyze the gathered data to offer insights into heat stress, changes in feeding efficiency, and the cow's oestrus. Special hormones are released when an oestrus cycle occurs, affecting the cow's behavior and movement (Roelofs et al. 2010).
- **AI in data collection:** Prior to now, gathered information was applied to a whole dairy farm. AI and other technology can supply unique data for each cow through the use of sensors, enabling farmers to make managerial decisions with more precision and accuracy (Tedeschi et al. 2021).

20.3.2 Managing Resources

By simulating human decision-making, AI enables producers to assess the data gathered by sensors and other hardware technologies and can offer interpretations and solutions, possibly changing how a dairy farm run (Niloofar et al. 2021).

Automation has assisted dairy farm owners in overcoming challenges in finding personnel willing to perform physical labor at dairy farms with the aid of lasers, sensors, and data collection. One benefit of automation is that animals appear to have adapted to this new technological trend and prefer robots to humans. Some of the managing resources details are as follows:

- **Drones:** Drone applications in the dairy sector are possible, but they frequently call for other technology. Drones can be used to check the herd or fences in general or to help herd cows from the fields to the barns. The addition of new technology opens up more possibilities. Visual sensors have shown to be useful for measuring pasture growth and doing land surveys. Drones are being used by modern dairy to map, examine, and take pictures of pastures to look for growth.

Drones can distinguish between cows and other animals, such as deer, thanks to algorithms. The likelihood of finding and following cows increases significantly when thermal imaging is combined with it, especially in fields with scattered trees or dense vegetation (Wich and Koh 2018). Farmers could spot anomalous behavior in the cow, such as lameness, illness, or calving, using temperature detection. Drones might be more practical in these applications if battery life is increased and autonomous flying capabilities are enhanced (Krishna et al. 2017).

- **Robots:** The most well-known application of robots in the dairy business is arguably robotic milking machines, which boost productivity and replace labor

that is either expensive or hard to find. In addition to reducing labor expenses, Lely's Astronaut A5 and DeLaval's Voluntary Milking System provide cows the freedom to choose when they want to be milked. Robotic milkers, or milkbots, autonomously locate the cow's teats, wipe the udders, and milk the animal (Kalantari et al. 2015).

Due to labor and time constraints, cows were routinely milked twice a day prior to the invention of robots. With the ability to milk cows up to three times a day, production and profits have been greatly increased. Additionally, there is a chance for medical and health evaluations using transponders or sensors while the cows are stationary for several minutes during milking. These devices can analyze the speed, volume, and quality of milk produced as well as how much the cow has eaten, its heat cycle, and other factors (Mottram 2016).

Robots might also be used to clean and sanitize the barn, improving biosecurity precautions and creating a healthier environment for the cows. Robots may possibly have a place in the calving process. For cows housed indoors, robotic aid may be available, albeit it would not be as helpful as it would be for a herd that is outdoors (Driessen et al. 2015).

- **3D Printing:** The dairy business has a wide range of 3D printing uses. Machine components are a common use for 3D printing, which may be particularly useful to farmers in rural areas since it can save them important time and even money, depending on the part required.
- **Augmented Reality:** The merging of digital data with the user's environment in real time is known as augmented reality (AR). According to studies, AR can be used to increase the visual attractiveness of food or accurately determine the right serving quantities.
- **Virtual Reality:** Virtual reality (VR) is referred to as a digital world that may be manipulated to appear real via the use of technology. Applications range in the dairy sector from farm tours to veterinary training, all of which have a favorable effect on efficiency and safety.

  The VR component is offered by a computer that enables students to see an object inside the cow realistically, allowing them to practice fertility checks like pregnancy detection or identify reproductive issues without putting the cow or the student in danger (Nayar 2004).
- **Block Chain**: It is common that customers are getting more and more interested in the origins and methods of food production. Blockchain can connect every link in the supply chain, from the producer to the consumer, enabling food safety and traceability.
- **Internet connection**: The Internet of Things (IoT) enables the linking of these technologies. For complete farm records, including field management, inventory, operations, grazing, and even biosecurity, a firm called AgriWebb uses IoT. IoT is used by Stellapps in India to provide a wide range of goods, including general herd management, milk evaluation, payment processing, and cold chain monitoring. Dell Technologies collaborates with the dairy manufacturer Chitale and is also significantly engaged in IoT applications.

All these technologies are opening up prospects for the dairy sector to boost productivity, profitability, and efficiency.

20.3.3 Livestock Welfare

The four-level IoTTs with AI and ML have the potential to manage the welfare of livestock. A track of animal activities and classification techniques, monitoring animal behavior at an individual level for decoding animal behavior, and examining of stress and edema in animals using an infrared thermal imaging techniques can be done using IoTTs. The attributes like appropriate behavior, good feeding, good health, good housing, and environmental conditions to assess the quality of the animal products received (Vanhonacker et al. 2010). The overuse of antibiotics and poor quality of animal feeds can also be linked to this perception. In the recent years, AI has significantly changed the veterinary sector by the following:

- Veterinary diagnostics easier
- Medical care accessible
- Data collection convenient

20.3.4 Monitoring Livestock

In general, cutting-edge technologies are excellent at deciphering several kinds of data, including text, audio, video, and photos. Then, inside such datasets, sophisticated algorithms can cluster, categorize, or forecast patterns. Advanced data analysis and algorithms have been used in animal production systems to use pattern recognition for disease detection and animal monitoring.

20.3.5 Diagnosing, Prevention, and Control of Livestock Disease

Identification, prevention, and prediction of diseases in animals is a big cost driver. In this case, modern technologies like AI, ML, sensors, and big data can be used. It helps in monitoring the health of the animals by keeping track of key health parameters: air quality, consumption of food and fluids, and movement. By the use of AI and ML techniques, disease outbreaks can be predicted and prevented by farmers. A system using such techniques is very beneficial: the production costs can be lowered as a small group of farmers can manage larger number of animals, the disease outbreaks leading to bigger loses can be prevented.

An outbreak of infectious disease can cause major loses of animals as these live together under one shelter. It becomes difficult to manage and contain a contagious disease if timely detection is not done and thus lead to severe financial loses. In turn, a farm with several smart sensors makes early detection of disease possible by analyzing the abnormal behavior of animal.

20.3.5.1 Body Condition Scoring

BCS helps in monitoring the health of individual cows and health status of herds. This score helps in changing the feeding patterns and reproduction management by analyzing the fat reserves in the cow. The applications like DenseNet, AlexNet, etc. are used for analyzing the datasets obtained by BCS. For example, the energy demand and temperatures go hand in hand which makes it necessary for adapting the feeding pattern according to the temperatures else it causes behavioral changes, lower BCS, and lower milk yields.

20.3.5.2 Metabolic Status

The metabolic disorders are seen in dairy cows during early lactation. The diagnosing and prevention of such disease before it causes major effect reduces costs of the treatment. Decision tree and random forest models are used for prediction of displaced abomasum and milk fever caused in dairy cows during early lactation.

20.3.5.3 Infectious Disease and Spatial Analysis

The spread of infectious diseases can be prevented using AI and ML approaches. The paratuberculosis disease in dairy cows is caused by *Mycobacterium* avium subspecies. The clustering methods were used to control and manage the spread of paratuberculosis. The hotspots of the disease outbreak could be managed using geospatial mapping techniques. Spatial analysis could help in limiting the spread of disease where rainfall and temperature were involved; for example, blue tongue disease caused by insects.

20.3.6 Livestock Management

The livestock management (LsM) involves management of two important resources, i.e., managing the quality of life of the livestock and managing the state of the land where the animals are being farmed. The quality-of-life management must include management of biosecurity, animal welfare, and product quality control. These play a major role in livestock management as these parameters will help in determining

the quality of life of animals and also shows their health and welfare status. Various techniques have been used in livestock management; for example, using IoT and deep learning has helped in effective management of livestock digitally, tracking of mating time for managing the time of oestrus for management of reproduction using IoT technologies, the livestock activity and location tracking using Iot-based LoRa network which uses machine-to-machine IoT protocol of connectivity which helps in transferring of data at a wide range.

The remotely controlled operations using advanced and enabled sensors and attenuators can be responsibly managed using the IoT model. This model makes use of protocol of stack data logging, transport management, processing and reporting and security, and privacy functionalities for positioning, touch and proximity and vision, etc.- based sensors.

20.4 Methodology of AI and ML

Currently, a number of industries are being transformed by technology including computers, sensors, cloud computing, ML, and AI. Greater gains and efficiencies are produced by them (Kunisch 2016). We must thus investigate how this cutting-edge technology might boost our animal production efficiencies and profits. Over the next three decades, there will be 70% growth in demand for different types of meat and animal products worldwide. In order to comprehend how various applications of animal farming technology might assist, farmers enhance animal health, boost revenues, and reduce their environmental impact (Neethirajan 2020). These methods and applications are as follows:

20.4.1 Models and Algorithms

In animal production, precision livestock farming techniques are the biggest help in increasing production and animal comfort. These techniques are focused on the areas of monitoring the animals' health, comfort, and production indications, such as body temperature for animals (Ferguson 2014). These temperatures provide information about animal health and productivity (Pomar 2019). There are two classical approaches that are used to estimate animal welfare:

1. Mechanistic modelling (MM)
2. Empirical modelling.

This modelling is basically based on the biophysical concept of energy conservation and mass in live animals. These models have the potential to solve complex problems like determining the optimal nutrient composition of animal feed and identifying functional limiting factors. To apply the mechanistic modes to animal farming, the collection of a large volume of diverse datasets is required. For

Fig. 20.5 The schema shows the different technologies that refers as advanced technologies, can help farmers to create better outcomes

example, local weather data, air quality data, voice signals of animals, animal behavior data, etc. There are various sensors that can help to capture this real-time data effectively. These approaches require the conceptualization of hypotheses centred on how a system works, and thus how variables are interconnected.

The following could serve as a summary of the general methodology used in the development of modelling:

1. Problem identification, hypothesis generation, and model bounds
2. Model conceptualization
3. Data collection
4. Model equation and assumptions
5. Model evaluation
6. Repeat steps 1 to 5 for the subsequent phase of model development (Fig. 20.5).

20.4.2 Sensors

In earlier times, farmers often handled the disease in their livestock by either doing nothing, using prescriptions from veterinary doctors or providing a mix of antibiotics. The invention of modern technologies like sensors provides enormous benefits for livestock farming farmers. These modern technologies can be used to identify solutions to problems in livestock farming. For example, it can improve the productivity and welfare of livestock by detecting sick animals and intelligently identifying areas for improvement. Instead of reacting to diseases after they become evident, the sensor provides an opportunity to constantly monitor key animal health parameters such as movement, air quality, food consumption, fluids, and sleep cycles.

A sensor can be defined as a device for measuring biological, chemical, physical, or mechanical properties, recording and collecting the data for interpretation by humans or machines. The sensor technologies can be classified into different types based on the requirements of the animal farming market. For example, sensors for precision milking robots and feeding systems are being used for poultry, swine, and other farm animals. Hardware sensors, such as cameras or vision sensors, temperature sensors, RFID tags, facial recognition machines, vision sensors, accelerometers, etc.

Farmers can easily predict and prevent abnormalities by monitoring their health with the help of collecting data on a regular basis. There are various sensors being used to classify the behaviors of animals, such as walking, grazing, ruminating, etc. There are also proven studies on different sensors like magnetometers (Dutta et al. 2015), optical sensors (Pegorini et al. 2015), or depth video cameras (Matthews et al. 2017), along with ML models that can help to classify and predict animal behavior.

20.5 Conclusion and Future Trends

The burgeoning human population and growing demand for animal food and products resulted in an urgent need to enhance livestock production. However, large-scale maintenance of farm animals, constant monitoring of their health, nutrition, and breeding and proper management of resources have become a challenge for stakeholders and farmers. Therefore, to overcome these difficulties and simplifying the otherwise laborious work resulting from traditional farming methods, researchers have started incorporating bigdata-based technologies such as AI and ML algorithms in order to enhance the livestock production and resource management. These technologies are completely based on the data collected via different sources, and further help farmers analyze the contemporaneous condition of farm animals, and also assist them by predicting the proper feed requirements, health status, and behavioral changes.

However, there are very few models available till date which can interconnect the datasets from different farms, resulting in limited use of this technology to varying environmental conditions. Moreover, there are less or no models available which combines the grazing and livestock health together or which can help analyze the comfort of animals under their environment except for certain temperature sensing-based model systems. Only two model systems available which can diagnose the disease or predict the treatment for farm animals that too based on just the lame behavior of animals.

Therefore, to encourage the application of these big data technologies in the livestock sectors, new models and innovative projects should be initiated, and new sensors should be proposed, in order to detect broad range factors so as to detect earlier disease symptoms.

References

Abraham M, Pingali P (2020) Transforming smallholder agriculture to achieve the SDGs. In: The role of smallholder farms in food and nutrition security. Springer, Cham, pp 173–209

Ayele J, Tolemariam T, Beyene A, Tadese DA, Tamiru M (2021) Assessment of livestock feed supply and demand concerning livestock productivity in Lalo Kile district of Kellem Wollega zone, Western Ethiopia. Heliyon 7(10):e08177. https://doi.org/10.1016/j.heliyon.2021.e08177

Bannink A, Van Schijndel MW, Dijkstra J (2011) A model of enteric fermentation in dairy cows to estimate methane emission for the Dutch National Inventory Report using the IPCC tier 3 approach. Anim Feed Sci Technol 166:603–618

Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7):711–720. https://doi.org/10.1109/34.598228

Borgonovo F, Ferrante V, Grilli G, Pascuzzo R, Vantini S, Guarino M (2020) A data-driven prediction method for an early warning of coccidiosis in intensive livestock systems: a preliminary study. Animals 10(4):747. https://doi.org/10.3390/ani10040747

da Rosa Righi R, Goldschmidt G, Kunst R, Deon C, da Costa CA (2020) Towards combining data prediction and internet of things to manage milk production on dairy cows. Comput Electron Agric 169:105156

Delgado C (2005) Rising demand for meat and milk in developing countries: implications for grasslands-based livestock production. In: McGilloway DA (ed) Grassland: a global resource. Academic Publishers, Wageningen, pp 29–39

Dora M, Kumar A, Mangla SK, Pant A, Kamal MM (2021) Critical success factors influencing adoption in food supply chains. Int J Prod Res 60(14):1–20

Driessen C, Heutinck LF (2015) Cows desiring to be milked? Milking robots and the co-evolution of ethics and technology on Dutch dairy farms. Agric Hum Values 32(1):3–20. https://doi.org/10.1007/s10460-014-9515-5

Dutta R, Smith D, Rawnsley R, Bishop-Hurley G, Hills J, Timms G, Henry D (2015) Dynamic cattle behavioural classification using supervised ensemble classifiers. Comput Electron Agric 111:18–28. https://doi.org/10.1016/j.compag.2014.12.002

Ellis JL, Jacobs M, Dijkstra J, van Laar H, Cant JP, Tulpan D, Ferguson N (2020) Synergy between mechanistic modelling and data-driven models for modern animal production systems in the era of big data. Animal 14(S2):s223–s237. https://doi.org/10.1017/S1751731120000312

Ezanno P, Picault S, Beaunée G, Bailly X, Muñoz F, Duboz R et al (2021) Research perspectives on animal health in the era of. Vet Res 52(1):1–15

Ferguson NS (2014) Optimization: a paradigm change in nutrition and economic solutions. Adv Pork Prod 25:121–127

Garcia R, Aguilar J, Toro M, Pinto A, Rodriguez P (2020) A systematic literature review on the use of machine learning in precision livestock farming. Comput Electron Agric 179:105826

Gertz M, Große-Butenuth K, Junge W, Maassen-Francke B, Renner C, Sparenberg H, Krieter J (2020) Using the XGBoost algorithm to classify neck and leg activity sensor data using on-farm health recordings for locomotor-associated diseases. Comput Electron Agric 173:105404

Haldar A, Mandal SN, Deb S, Roy R, Laishram M (2022) Agriculture, livestock production, and aquaculture: advances for smallholder farming system. In: Agriculture, livestock production and aquaculture. Springer, Cham, pp 173–218

Hidalgo A, Zouari F, Knijn H, van der Beek S (2018) Prediction of postpartum diseases of dairy cattle using macehine learning. In: Proceedings of the world congress on genetics applied to livestock production. World congress on genetics applied to livestock production, Rotterdam, p 104

Kalantari AS (2015) Using mathematical modeling techniques for optimized dairy herd management and decision making. The University of Wisconsin-Madison, Madison

Kebreab E, Smith T, Tanner J, Osuji P (2005) Review of undernutrition in smallholder ruminant production systems in the tropics. In: Ayantunde AA, Fernandez-Rivera S, McCrabb G (eds) Coping with feed scarcity in smallholder livestock systems in developing countries, pp 3–94

Krishna KR (2017) Push button agriculture: robotics, drones, satellite-guided soil and crop management. CRC Press, Boca Raton

Kumar A, Singh KM, & Singh R (2012) Role of livestock sector in sustainable livelihood security in Bihar: status and opportunities. Available at SSRN 2062823. https://doi.org/10.2139/ssrn.2062823

Kunisch M (2016) Big data in agriculture–perspectives for a service organization. Landtechnik 71(1):1–3

Kunkel KR (2003) Facpptory farming as a social problem. Social problems*:* constructionist readings. Taylor Francis, Routedge, pp 101–106

Linaza MT, Posada J, Bund J, Eisert P, Quartulli M, Döllner J et al (2021) Data-driven applications for sustainable precision agriculture. Agronomy 11(6):1227. https://doi.org/10.3390/agronomy11061227

Liu C, Zoph B, Neumann M, Shlens J, Hua W, Li LJ et al (2018) Progressive neural architecture search. In: Proceedings of the European conference on computer vision (ECCV). Springer, New York, pp 19–34

Matthews SG, Miller AL, PlÖtz T, Kyriazakis I (2017) Automated tracking to measure behavioural changes in pigs for health and welfare monitoring. Sci Rep 7(1):1–12

Mohamed ES, Belal AA, Abd-Elmabod SK, El-Shirbeny MA, Gad A, Zahran MB (2021) Smart farming for improving agricultural management. Egypt J Remote Sens Space Sci 24(3):971–981

Mottram T (2016) Animal board invited review: precision livestock farming for dairy cows with a focus on oestrus detection. Animal 10(10):1575–1584. https://doi.org/10.1017/S1751731115002517

National Research Council (2003) Air emissions from animal feeding operations: current knowledge, future needs. National Academic Press, Washington

Nayar PK (2004) Virtual worlds: culture and politics in the age of cyber technology. SAGE Publications India, Delhi

Neethirajan S (2020) The role of sensors, big data and machine learning in modern animal farming. Sens Bio-Sens Res 29:100367

Nikoloski S, Murphy P, Kocev D, Džeroski S, Wall DP (2019) Using machine learning to estimate herbage production and nutrient uptake on Irish dairy farms. J Dairy Sci 102(11):10639–10656. https://doi.org/10.3168/jds.2019-16575

Niloofar P, Francis DP, Lazarova-Molnar S, Vulpe A, Vochin MC, Suciu G et al (2021) Data-driven decision support in livestock farming for improved animal health, welfare and greenhouse gas emissions: overview and challenges. Comput Electron Agric 190:106406

Parkhi OM, Vedaldi A, Zisserman A (2015) Deep face recognition. In: Proceedings of the British Machine Vision Conference (BMVC). BMVA Press, Durham, pp 41.1–41.12

Pegorini V, Zen Karam L, Pitta CSR, Cardoso R, Da Silva JCC, Kalinowski HJ, Ribeiro R, Bertotti FL, Assmann TS (2015) In Vivo Pattern Classification of Ingestive Behavior in Ruminants Using FBG Sensors and Machine Learning. Sensors. 15(11):28456–28471. https://doi.org/10.3390/s151128456

Piles M, Fernandez-Lozano C, Velasco-Galilea M, González-Rodríguez O, Sánchez JP, Torrallardona D et al (2019) Machine learning applied to transcriptomic data to identify genes associated with feed efficiency in pigs. Genet Sel Evol 51(1):1–15

Pomar C, Remus A (2019) Precision pig feeding: a breakthrough toward sustainability. Anim Front 9(2):52–59. https://doi.org/10.1093/af/vfz006

Pomar J, López V, Pomar C (2011) Agent-based simulation framework for virtual prototyping of advanced livestock precision feeding systems. Comput Electron Agric 78(1):88–97. https://doi.org/10.1016/j.compag.2011.06.004

Roelofs J, Lopez-Gatius F, Hunter RHF, Van Eerdenburg FJCM, Hanzen CH (2010) When is a cow in estrus? Clinical and practical aspects. Theriogenology 74(3):327–344. https://doi.org/10.1016/j.theriogenology.2010.02.016

Sadeghi M, Banakar A, Khazaee M, Soleimani MR (2015) An intelligent procedure for the detection and classification of chickens infected by clostridium perfringens based on their vocalization. Braz J Poultry Sci 17:537–544

Tedeschi LO, Greenwood PL, Halachmi I (2021) Advancements in sensor technology and decision support intelligent tools to assist smart livestock farming. J Anim Sci 99(2):skab 038

Thornton PK (2010) Livestock production: recent trends, future prospects. Philos Trans R Soc B: Biol Sci 365(1554):2853–2867

Umutoni C, Augustine A, Germain JS (2015) Evaluation of feed resources in mixed crop-livestock systems in Sudano-Sahelian zone of Mali in West Africa. Int J Livestock Res 5(8):27–36. https://doi.org/10.5455/ijlr.20150813090546

Vanhonacker F, Van Poucke E, Tuyttens F, Verbeke W (2010) Citizens' views on farm animal welfare and related information provision: exploratory insights from Flanders, Belgium. J Agric Environ Ethics 23(6):551–569. https://doi.org/10.1007/s10806-010-9235-9

Wich SA, Koh LP (2018) Conservation drones: mapping and monitoring biodiversity. Oxford University Press, Oxford

Zhang C, Xiao D, Yang Q, Wen Z, Lv L (2020) Application of infrared thermography in livestock monitoring. Trans ASABE 63(2):389–399. https://doi.org/10.13031/trans.13068

Chapter 21
Applying Sensors and Robotics Towards Smart Animal Management

Neeraj Kashyap and Bharti Deshmukh

Abstract Increasing the scale, and scope of livestock farming as a primary commercial occupation on an industrial scale, has required the application of management principles ensuring maximization of the profits while optimizing the uses of available resources. The success of livestock farming is closely related to understanding animal biology, and deploying that understanding towards minimization of inputs, optimization of production, and maximization of produce quality by precise management of farm operations. The farm decision-making, which required engaging skilled personnel at a fair cost earlier, can be reduced to a large extent by the deployment of smart farm management with the recent technical advances. Smart farm management is aided by data logging tools equipped with sensors, and automation tools based on robotics, connected to some data-driven farm decision (DDFD) system. The observation of intricate fluctuations in the biorhythms of the animals as an indicator has opened up a possibility of early and accurate prediction of events of importance towards animal health, reproduction, production, and welfare while reducing the human error in observation. Development, and affordability of various sensors that can be used to monitor location, movement, sound, physiological parameters, product quality, etc., and their integration with the development of animal-friendly wearable technologies, can generate a tremendous amount of data by continuous real-time data logging of individual animals. This big data can fuel the farm decision called DDFD. This DDFD is further utilized to automate certain farm operations with robotics. Smart animal management is an integration of all such technologies to achieve precision animal farming.

N. Kashyap (✉)
Department of Bioinformatics, College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University (GADVASU), Ludhiana, Punjab, India
e-mail: neerajkashyap@gadvasu.in

B. Deshmukh
Present Address: Department of Animal Genetics, and Breeding, College of Veterinary Sciences, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_21

Keywords Sensors · Robotics · Smart animal management · Data-driven farm decision

21.1 Introduction

Since the early civilization days, animals have been playing a profound and integral role in human lives in social, cultural, economic, and nutritional aspects. The human-animal bonding started from the dogs, which have been human companions for thousands of years, to assist in guarding, tracking, and rescue operations, helping the disabled, besides being a companion. Human society is benefitted by domesticating animals in aspects of labor, nutrition, transportation, etc., such that they were seen as a form of wealth by many civilizations, and termed by the word "livestock". Even though the industrialization through the nineteenth century has limited the use of livestock for labor, and transport, the overall scale of livestock farming has expanded, due to the increased demand for animal source food in the past century. With the increasing animal product demand, livestock farming is gradually shifting from low-input small ventures to medium- and large-scale intensive farms. Farmers nowadays have been gradually adapting livestock farming as a primary commercial venture rather than just a supporting farming activity. The size and scope of operations at these livestock farms are also expanding, given facts of ensuring maximization of profits from these farms. Profit maximization requires us to adopt strategies for efficient farm management, minimize the duration for which animals are raised as unproductive, reduce the animal losses due to disease, disability, or death, optimize animal reproduction, and ensure optimum nutrition.

Achieving the state of minimal waste/losses and maximum utilization at the farms entails precise and prompt farm decisions. Such decisions require the employment of skilled farm personnel, which in itself is a costly affair in long term. However, the minimization of human intervention by utilization of modern techniques for animal monitoring, decision support, and automation of farm operations has been made possible and is gradually being affordable by technological advances for deploying sensors, and robotics for livestock farm operations.

21.2 Challenges to Traditional Livestock Farms

Traditional livestock farming mainly involved managing animals small scale, mostly secondary to crops. These small farms had the personal involvement of the farmer for individual care of the animals. With the increasing human population, and improving purchase ability of the people, the demand for animal products has shown an increasing trend. To meet the demand, livestock farming must evolve to upscale production while improving efficiency. The concerns about animal welfare, public health, and environmental sustainability also need to be addressed. With the intensification, commercialization, and industrialization of livestock farming, the size of these farms increased. The major challenges in effectively monitoring

animals in such farms are cost, accuracy, and promptness of reports. The majority of previously used methods for monitoring livestock are time-consuming, labor-intensive, and therefore costly (Nasirahmadi et al. 2017). To detect animal health, and production-related issues at farms, the farmers have to mostly rely upon observations from stock people (animal handlers employed at the farms). Traditionally, the only available option to ensure maximum coverage of reporting such issues was to employ a more skilled stock person and to increase stock person availability. However, it remains a fact that even an attentive, and skilled stock person might miss animals in crucial points, due to non-continuous monitoring, and non-overt signs. Thus, to fill in the gap, livestock farming has been aided with third-party auditing programs, which offer comprehensive animal farm assessments, but these, too, often are time and cost demands. The monitoring criteria applied are mostly subjective measures like detection of body condition score, lameness, and lesions. Yet, there is a problem of inconsistency with these auditors. The more accurate auditors are mostly invasive, and they typically require restraining the animals by stock people, which causes additional stress to the animals.

To enhance the overall efficiency of farm operations, and to save labor, the precision livestock farming (PLF): A concept of integrating information, and communication technologies (ICT), has emerged in recent years to aid in farm decisions. The Internet of Things (IoT) devices has paved a wire-free, and system-independent approach for animal monitoring. PLF uses ICT and IoT tools to improve and automate monitoring data generation on animal production, reproduction, health, and environmental impact through various technologies such as sensors viz. cameras, microphones, tags, and deployment of intelligent software for data-driven farm decisions (DDFD). PLF technologies can automate livestock farming by allowing farmers to monitor various parameters for the health and welfare of a large population of animals. They can detect instances of health/welfare importance in individual animals promptly, and may even give an anticipatory alarm based on previous data (Benjamin and Yik 2019). Such use of advancing technologies will help secure a better livestock health and welfare status.

21.3 Integration of Technology in Livestock Farming

While it is certain that the technology is setting its importance in livestock farming systems, and is proving to be economically beneficial in long-term implementation, it remains the choice of the farmer to decide the level of technological interventions to be applied in livestock farming. As such, there are three tiers of technology applications at the farms, namely, PLF, digital farming, and smart farming.

21.3.1 Precision Livestock Farming

Precision livestock farming, abbreviated as PLF, can be explained as the automated identification and monitoring of individual animals for capturing data on their performance, fitness, and welfare through parameters such as temperature, motion, images, sounds, location, weight, and biological metrics in livestock analyzed in real-time (Berckmans 2014; Neethirajan 2017). PLF involves meticulously controlled, accurate, and optimized livestock production to simultaneously facilitate efficient resource utilization, improved yield, reduced risk, and minimal environmental impact. The availability and scope of PLF have rapidly expanded due to progress in computation, and the advent of affordable sensors combined with the connectivity for capturing, and processing data.

The main objectives of PLF are to optimize livestock feeding, mitigate environmental stress, manage crops in perfect synergy with livestock, ensure food safety through traceability of products, and improve animal health, and crop efficiency (Pomar et al. 2011). To improve farming efficiency, the data continuously generated in livestock farms need to be managed (Suryawanshi et al. 2017). The data management, if done correctly, can result in improved productivity, pasture management, livestock nutrition, and animal health.

21.3.2 Smart Livestock Farming

The distinction between smart farming from precision farming lies in the fact that smart farming does not necessitate precise measurements, rather it focuses on capturing data and interpreting them using computing tools to make farm operations more predictable, and efficient. So, smart farming is essentially an integration of PLF with smart software driven through systems' capability of decision making based on data inputs from ICT-based platforms. Additionally, smart farming comprehensively integrates farming operations, labor management, personnel deployment, risk management, maintenance, yield calculation, purchases, warehousing, logistics, and marketing into a single system.

21.3.3 Digital Livestock Farming

Digital livestock farming essentially implies creating valuable information from livestock farming data. Digital farming ventures go ahead of the presence, and availability of data to utilize, and develop actionably, and timed insights into the data originating from livestock farms using PLF or traditional tools. Digital farming in its essence is an integration of both concepts—precision farming and smart farming. Digital livestock farming is now being viewed as the future of the farming industry.

- **Continuous individual monitoring:** With an increasing number of heads of livestock per stock person, and increasing relevance of welfare, one has to rely on technology to ensure animals are monitored 24 h a day. Additionally, many criteria for optimal welfare are in line with production criteria (better nutrition, health care, stress minimization, etc.). This continuous monitoring mainly focuses on daily time budget and biorhythms.
- **Daily time budget:** Every animal shows a pattern of duration spent for eating, drinking, resting, interacting, etc. in accordance to their age, physiological state, management, and other factors, termed as "daily time budget". A significant departure of an animal or a group of animals from its "daily time budget" may indicate the presence of a problem or a change in status.
- **Biorhythms:** The bio-physiological parameters like pulse rate, respiration rate, blood pressure, rumination, core body temperature, etc. exhibit cyclic changes over the duration, called "biorhythm". Any departure from the routine biorhythm indicates the presence of a problem or a change in the bio-physiological status of the animal.

21.4 Tools of Smart Livestock Farming

In addition to precision, and efficiency, the optimization of processes, and anticipatory planning are key aspects of smart farming. Fundamental technologies for smart farming are tracking systems, such as GPS. They enable data to be allocated to a particular region of the farmland or determine the current position of agricultural machines or animals in the barn. However, in the long run, the effective, and useful implementation of big data for smart farming in the future requires the development of a nationwide digital infrastructure, especially in rural areas.

The widely used components that ICT-based platforms use for smart farming are smartphones, and tablets, wearable technologies, IoT, robotics, drones, etc. equipped with various sensors, and actuators driven and coordinated through Artificial Intelligence (AI), and Cloud Computing (CC).

21.4.1 Sensors

An understanding of biometric sensors, the algorithm of data logging, and machine learning is required for grasping the utilization of such sensors for improvement in livestock farming practices from traditional to digital. Quite a lot of sensors viz. RFID tags, accelerometers, microphones, cameras, thermistors, etc. are nowadays used on the farms as installable and wearable technologies to monitor and capture information from individual animals or herds. The data captured from these sensors are stored locally or over the cloud and processed using algorithms. An algorithm can be explained as a step-wise set of instructions to execute a chain of operations, purposed to address a problem or to automate a task. An algorithm's ability to

associate the data captured by a sensor to a biological outcome with desirable accuracy is the deciding factor of its relevance to a farmer. As an example, a variable "unsteady gait" of an animal captured by pedometers over time can be used to detect "lameness" or the accelerometer measured changes in rumination frequency can detect the biological outcome "stress".

A family of programming computational techniques that uses large sets of example data to tune and configure its algorithm is referred to as "machine learning". A machine learning system can become proficient at processing and analyzing big data sets to track variables and produce estimates at a rate that is not achievable by manual or traditional statistical methods, due to its inherent learning capability using the training data (Puri et al. 2016). The working pipeline can be summarized as the data from remote monitoring sensors related to animal identification, over-time observations, and production that are fed to trained algorithms which then provide timely inferences and alarms relevant to the health, productivity, and welfare status of livestock.

21.4.1.1 Radio Frequency Identification (RFID) Tags

The use of the RFID chip is a well-established technology for livestock identification for any purpose on farms. These tags are devices having transponder circuits, that can emit or reflect radio waves (at a low, high, or ultra-high frequency) to communicate with the RFID reader to wirelessly read, and write data. The RFID tags store information primarily on animal identification and are primarily implanted as ear tags. These RFID tags when coming inside the range of any RFID reader can emit and receive signals; the tag receives a signal from the RFID reader, and produces a second radio frequency signal in response, carrying data that is then received by the RFID reader (Ariff et al. 2014).

Low-frequency (LF) RFID has low distance readability and is thus a valuable component of farm automation where the proximity of an animal needs to trigger an action such as automated feed dispensers. These RFID readers may be fitted with an enclosed feeder that allows one RFID-tagged animal at a time, and the system can then dispense a limited diet to the animal, specific to its requirement, taking into account the identification, and feeding frequency of the animal via RFID (Maselyne et al. 2015). They may also be fitted at milking parlors to automate milk yield recording. Changes in the frequency of visits of certain sows could be noticed to indicate health issues or abnormal behaviors to be addressed early (Cornou et al. 2008). These LF-RFIDs have a low read range of lesser than 1 m and identify only one animal at a time. Ultra-high frequency (UHF) readers can be used to track multiple animals at a greater range (Adrion et al. 2017).

21.4.1.2 Cameras

The cameras are valuable sensors to collect the image and video footage of animals. The images may further be either flat (2D) depth enabled (3D) or temperature-

sensitive (infrared). The techniques like computer vision and image analysis can be used to translate the captured images into indicators of animal distribution, and activity such as animal location, proximity, position, and movement (Kashiha et al. 2014). Imaging in animals can also be used to estimate weight from body area, behavior, movement patterns, etc. The 2D camera sensors require sufficient lighting and a contrasting background, which is not always possible in animal housings. This limitation can be addressed by the use of the 3D camera (three-dimensional depth-based sensors) such as Intel RealSense and Microsoft Kinect cameras. These high-definition cameras are equipped with an infrared illuminator and depth sensors. Infrared imaging is of high merit during observing nocturnal behaviors, detecting surface temperature gradients to detect injury/inflammation, while depth sensors can determine the proximity of the animal to the objects.

The fact that any object having a temperature above absolute zero emits infrared radiation, with a proportional increase in wavelength of emitted radiation with the temperature of the object is used as the working principle of infrared imaging. Thermal imaging captures the infrared radiation and converts it into an image with graded intensity. The infrared cameras are used to non-invasively measure the real-time temperature of the animal body surface and its physiological and pathological temperature alterations (Salles et al. 2016).

21.4.1.3 Microphones

The microphones can convert sound waves into electrical signals that may be utilized by algorithms with the intent to detect, classify, and localize specific sounds such as vocalization patterns or respiratory noise as indications of stress or illness (Schön et al. 2001). For instance, change in wavelength, pitch, and frequency of vocalization can be associated with stressful situations, oestrus, or pain (Yoshihara and Oya 2021). Therefore, microphones and vocalizations could become an automatic daily measure.

Vocal expressions of animals change as a result of emotional statuses in specific situations. Therefore, the distress calls of pigs can be used as indicators of impaired welfare. Manteuffel et al. (2004) and Schön et al. (2004) introduced vocalization in livestock farms as a measurement of welfare. They began by classifying pig calls as either contact calls (grunts) or calls reflecting arousal (squeals and screams). Later on, the same team developed STREMODO, the first system that can be used in housing, during transport, and in abattoirs, to identify stress vocalizations.

21.4.1.4 Thermistors

Thermistor is a sensor that can monitor the temperature of a contact media. The biometric monitoring devices for animals mostly utilize thermistors embedded in a data logger or wearable tags. To take temperature measurements, the sensor requires to be in direct contact with the body. The non-invasive thermistors can thus only record the peripheral temperature. The peripheral temperature readings depend not

only on core temperature but also largely influenced by environmental conditions, exposure to solar radiation, and peripheral blood circulation. The core body temperature can be logged by a thermistor embedded invasively inside the body, mostly in the muscles, rumen, or vagina. The core body temperature shows variation in response to disease and oestrus, and thus can be used as an indicator of health, stress, and reproduction. Thermistors can measure temperature with accuracies up to ±0.1 °C (Sellier et al. 2014).

21.4.1.5 Accelerometers

Any kind of movement can be detected by an accelerometer, a lightweight and small sensor. The working principle of an accelerometer is based on measuring accelerating forces by capturing the vibrations and converting them into piezoelectric effects using microscopic crystals. The sensor determines the pace and direction of the movement by interpreting the amount of generated voltage. Forces can be static like gravity or acceleration due to movement. An accelerometer is one of the most applicable sensors for monitoring livestock behavior which can be integrated into hordes of wearable devices and record animal movement, mastication, ruminal motility, tail movements, etc. Inertial measurement units (IMUs) consisting of three-axis accelerometers and gyroscopes can be used to measure the precise movement trajectory provided sufficient sample rates are used. Unlike accelerometers alone, IMUs can measure both linear and angular acceleration.

Leg-mounted accelerometers, commonly referred to as pedometers, were the first applications used with cattle; and they are commonly used to measure dairy cows' lying time, and walking with accuracy (Alsaaod et al. 2015). Accelerometers have been primarily used in the dairy industry to measure activity related to oestrus behavior as the activity pattern changes dramatically during oestrus. The use of the accelerometer to measure feeding and rumination time has gone to a commercial level. Accelerometers have also been used for the evaluation of steadiness in gait to detect lameness (Chapinal et al. 2011).

21.4.1.6 Positioning Systems

Positioning systems are used to locate the animals and location is used in integration with farm management. Much valuable information can be deciphered from the animal's location, and its derivatives like velocity and direction. The livestock positioning can be adapted in two different ways such as local and global.

- **Local positioning:** Several techniques can be used to locate animals on the farm premises such as radar technology (Gygax et al. 2007) and Bluetooth to track the position of cows in a barn (Bloch, and Pastell 2020). Several indoor positioning systems are commercially available, targeting livestock farms particularly. These systems use a signal transmitting tag that can be used to trace the location of the

animal in real-time. These positioning data can be then used to infer about animals' behavior, health, and reproductive events such as feeding time, social interactions, lameness, oestrus, and many more.

- **Global positioning:** Satellites can be used to track objects where satellites can have an unobstructed view. The United States NAVSTAR Global Positioning System (GPS) and the Russian GLONASS are the current existing positioning systems. So far, GPS has been used mainly to study grazing behavior (Sickel et al. 2004) and to identify walking, resting, and eating behavior, when combined with accelerometer measurements (Williams et al. 2016). These systems are of particular use in free-range farming systems, where manually locating an animal is of importance.

21.4.1.7 Heart Rate

Cardiovascular system function and cardiac autonomic modulation are reflected by heart rate and heart rate variability. Since heart rate shows the immediate response to various stressors viz. exercise, pain, emotions, disease, etc., this information can be used to estimate physiological and psychological stressors in animals. Electrocardiography, the process of recording the electrical activity of the heart, is rather a sophisticated technique to be applied to farm animals for continuous heart rate monitoring. Rather, optical methods for measuring heart rate have cost-effective applicability. In cattle, a photonic sensor measuring the movement of the skin surface can be used to measure the cow's heart, respiration rates, and chewing patterns using a contactless setup (Beiderman et al. 2014).

21.4.2 Robotics

Digitalization and automation are expanding into many areas, resulting in the more widespread use of partially, and fully autonomous machines and robots. Automation has been the first line of development in any field including the livestock sector. In the livestock farming, automation solutions currently are successfully developed for various commercial applications such as environmental control, animal/herd management, milking (Automatic Milking Systems [AMS]), feeding (Automatic concentrate dispenser and Automatic Feeding Systems), and manure management (Automatic Robotic Systems) through robots or intelligent machines able to interact with their work environment and, in most of the cases, without a direct human control.

21.4.2.1 The Automatic Milking System

AMS is a classical farm automation solution that has revolutionized dairy farming around the globe. The world's first commercial robotic milking rotary has been unveiled by Swedish dairy equipment company 'DeLaval' at a pilot farm at Quamby Brook, Tasmania, Australia. Featuring five robots, the rotary can milk up to 90 cows per hour (Legg 1993). The term "AMS" refers to a system that automates all the functions of the milking process, and cow management is undertaken in conventional milking by a mix of manual and machine systems (De Koning and Rodenburg 2004). With the automation of the process, milking is no longer required to be performed at definite times; rather, the cow can now choose when to be milked in AMS, throughout a period of 24 h. For robotic milking farms, simulation optimization based on animal behavior was developed earlier in this century.

21.4.2.2 Robotic Feed Dispensers

The auto-feed dispenser has been developed to be used for pet animals. The companion animals like cats, and dogs besides hamsters, guinea pigs, and fish, live in the household and are partially or completely restricted in movement. Except for the cats, which can care for themselves, other animals like dogs are to a large extent dependent, while fish is completely dependent on humans. Automatic feeders provide a basic food supply. There are simple versions for pets—feeding robots, on the other hand, are found mainly in stables.

A microchip feeder by 'Sure Petcare' (https://www.surepetcare.com) has been designed for multi-pet households, which ensures animal-specific food dispensing. An automatic closing lid ensures that food stays fresh longer. A Wi-Fi-enabled automatic pet feeder 'Feeder robot' (https://www.litter-robot.com/feeder-robot.html) can be operated through unit control or an app 'AutoPets Connect'. It offers customizable programs to cater mealtime needs of the pets even when the owner is not at home (Bendel 2022).

In dairy farms, similar computer-controlled calf feeders have been deployed with many advantages over traditional calf feeding methods. It is possible to control and monitor the daily intake of individual calves equipped with a transponder (Reinemann and Helgren 2004). These calves learn to use the automated milk feeding system fairly easily, and a significant reduction in labor cost (73%) is achievable in calf care.

The belt feeder type feed distributor is a practical concept of an automatic feed supply system, that can be introduced to the farms. Optimization of the supply of feed with the exact ration at the right time for each animal can be the automated feed dispensing system. The automatic feeding system (AFS) by Lely Vector system consists of three main units: (1) a feed repository, an enclosed feed storage area; (2) a feed grabber with a bridge crane that can load roughages into the mixing bin; and (3) mixing and feeding robot (MFR) that can automatically mix and distribute feed to

the animals. Tangorra and Calcante (2018) could demonstrate that adapting AFS can reduce energy consumption, and labor by 97% and 79%, respectively, compared to a conventional approach (tractor + trailed TMR wagon), reducing the daily cost of feeding TMR up to 33% in dairy farms.

21.4.2.3 Dung-Removal Robot

The classical dung scrappers have been successfully used inside the animal sheds with the option of automated running at scheduled timings. Ahead of them, there are battery-powered dung-removal robots such as Lely Discovery Mobile Barn Cleaner 90 SW (Lely Industries, Maaslouis, Netherlands). It can clean perforated floors by scraping the manure and pushing it through the openings in the floor. The device also has a spray function (Leinweber et al. 2019).

21.4.2.4 Robotic Shearing

According to a 2016 report from the Australian Bureau of Statistics, there are about 2800 shearers, and 73 million sheep in Australia, depicting a national shortage of shearers, despite the rising global demand for wool, and record-high prices per kilogram. To address that, the Australian Wool Institute in partnership with Ranken Research, and Robo Shear, started to develop a prototype machine for fully automated end-to-end wool harvesting in 2019.

21.4.2.5 Robotic Egg Collectors

With the shift of public opinion on the cage system of layer rearing as cruelty and subsequent prohibition of the same by law in many countries, farmers are now shifting towards deep litter/barn system/ free-range housing. Under the cage systems, egg collection is not a problem; however, the same cannot be said about open houses, where eggs are laid around randomly. Furthermore, the egg is laid throughout the day, interfering with other farm activities, and requiring extra laborers. Various egg collection robots are in development. The robotic egg collectors use an array of sensors to find and collect eggs without disturbing the chickens. A commercially available robot 'Izbot' by Izario Poultry Robotics operates continuously throughout the production cycle in broiler breeder, and commercial egg-laying farms. The robot is capable of carrying out farm operations such as floor egg retrieval, bird health, and welfare monitoring along with monitoring the shed's environmental output (https://www.izario.co/).

21.5 Challenges in Implementing New Technologies

It is to be considered that not all technologies being developed will become practical on-farm products. Technology must be practical, feasible, affordable, cost-effective, and functional, besides addressing a specific need and being easy to operate, to be successfully adopted by the farmers. The reliability and durability of the technology decide the longevity of the technology considering the farm working conditions. The equipment must allow for proper washing and disinfection. While developing technology, it is essential to consider the need of the farmers while integrating the experience of farmers for solving implementation problems.

The need for internet access, which is essential for IoT, and cloud-connected smart devices, may be limiting for rural farms. Whereas some of the sensors may be liable to breaking down or corruption by exposure to physical force or humidity. The consideration of the safety of the animals while installing such devices on farms/animals must be also considered. Most of the sensors will require power to run, and the need to connect equipment to the power supply to charge the batteries can pose a hurdle to implementation.

21.6 Conclusion

Implementation of the technological advances in sensors and robotics at the animal farm premises are simultaneously benefitting the farmer, the animal, the consumer, and the science on different fronts by generating and utilizing hordes of data. The sensors and robotics are being developed to help the farmers, by reducing their labor requirement, improving monitoring efficiency, assisting in decision making, and ensuring timely operations to minimize the losses associated with mismanagement, thereby minimizing the risks associated with the livestock farming. It is doubtful though, whether these technologies can replace the experience, and knowledge of farmers, but it is certain to help them. The animals' welfare is also improved to a great extent by such technologies as they are ensured to get timely attention towards their health, discomfort, and cleanliness altogether with proper nutrition, and time management. Consumers are now being more and more aware of the source, and quality of the animal source food for health, and humanitarian concerns. They want to ensure wholesome animal products from the animals with ensured welfare, to not feel bad or guilty while consuming animal products. Denouncing the cage layer system is one such example.

The science is benefitted from the high throughput phenomics data being generated from such sensors, which in turn provides the raw material for a better understanding of animal biology and production science. These high quality and quantity of phonemic data are also the foundation of a successful breeding program for wholesome, and correctly directed genetic improvement of the animals, ultimately benefitting farmers, animals, and society. Last but not least, there is a

substantial initial investment involved in installing the new technology on the farm, which is out of the reach of most of the farmers in developing countries.

References

Adrion F, Kapun A, Holland EM, Staiger M, Löb P, Gallmann E (2017) Novel approach to determine the influence of pig, and cattle ears on the performance of passive UHF-RFID ear tags. Comput Electron Agric 140:168–179. https://doi.org/10.1016/j.compag.2017.06.004

Alsaaod M, Niederhauser JJ, Beer G, Zehner N, Schuepbach-Regula G, Steiner A (2015) Development, and validation of a novel pedometer algorithm to quantify extended characteristics of the locomotor behavior of dairy cows. J Dairy Sci 98:6236–6242. https://doi.org/10.3168/jds.2015-9657

Ariff MH, Ismarani I, Shamsuddin N (2014) RFID based systematic livestock health management system. In: 2014 IEEE conference on systems, process, and control (ICSPC 2014), pp 111–116

Beiderman Y, Kunin M, Kolberg E, Halachmi I, Abramov B, Amsalem R, Zalevsky Z (2014) Automatic solution for detection, identification, and biomedical monitoring of a cow using remote sensing for optimised treatment of cattle. J Agri Eng 45(4):153–160. https://doi.org/10.4081/jae.2014.418

Bendel O (2022) Passive, active, and proactive systems, and machines for the protection, and preservation of animals, and animal species. Front Animal Sci 3:1–7. https://doi.org/10.3389/fanim.2022.834634

Benjamin M, Yik S (2019) Precision livestock farming in swine welfare: a review for swine practitioners. Animals 9(4):133. https://doi.org/10.3390/ani9040133

Berckmans D (2014) Precision livestock farming technologies for welfare management in intensive livestock systems. Rev Sci Tech 33(1):189–196. https://doi.org/10.20506/rst.33.1.2273

Bloch V, Pastell M (2020). Monitoring of cow location in a barn by an open-source, low-cost, low-energy bluetooth tag system. Sens, 20(14), 3841. https://doi.org/10.3390/s20143841

Chapinal N, Passille AM, Pastell M, Hänninen L, Munksgaard L, Rushen J (2011) Measurement of acceleration while walking as an automated method for gait assessment in dairy cattle. J Dairy Sci 94:2895–2901. https://doi.org/10.3168/jds.2010-3882

Cornou C, Vinther J, Kristensen AR (2008) Automatic detection of oestrus, and health disorders using data from electronic sow feeders. Livest Sci 118(3):262–271. https://doi.org/10.1016/j.livsci.2008.02.004

De Koning K, Rodenburg J (2004) Automatic milking: state of the art in Europe, and North America. Automatic milking: a better understanding. wageningenacademic, wageningen, pp 27–37

Gygax L, Neisen G, Bollhalder H (2007) Accuracy, and validation of a radar-based automatic local position measurement system for tracking dairy cows in free-stall barns. Comput Electron Agric 56:23–33. https://doi.org/10.1016/j.compag.2006.12.004

Kashiha MA, Bahr C, Ott S, Moons CP, Niewold TA, Tuyttens F, Berckmans D (2014) Automatic monitoring of pig locomotion using image analysis. Livest Sci 159:141–148. https://doi.org/10.1016/j.livsci.2013.11.007

Legg B (1993) Hi-tech agricultural engineering–a contradiction in terms or the way forward. Mech Inc Eng 5(4):86–90

Leinweber T, Zähner M, Schrade S (2019) Evaluation of a dung-removal robot for use in dairy housing from an ethological, and process-engineering point of view. Landtechnik 74(3):55–68

Manteuffel G, Puppe B, Schön PC (2004) Vocalization of farm animals as a measure of welfare. Appl Anim Behav Sci 88:163–182. https://doi.org/10.1016/j.applanim.2004.02.012

Maselyne J, Saeys W, Van Nuffel A (2015) Quantifying animal feeding behaviour with a focus on pigs. Physiol Behav 138:37–51. https://doi.org/10.1016/j.physbeh.2014.09.012

Nasirahmadi A, Edwards SA, Matheson SM, Sturm B (2017) Using automated image analysis in pig behavioural research: assessment of the influence of enrichment substrate provision on lying behaviour. Appl Anim Behav Sci 196:30–35. https://doi.org/10.1016/j.applanim.2017.06.015

Neethirajan S (2017) Recent advances in wearable sensors for animal health management. Sens Bio-Sens Res 12:15–29. https://doi.org/10.1016/j.sbsr.2016.11.004

Pomar J, Lopez V, Pomar C (2011) Agent-based simulation framework for virtual prototyping of advanced livestock precision feeding systems. Comput Electron Agric 78(1):88–97. https://doi.org/10.1016/j.compag.2011.06.004

Puri M, Solanki A, Padawer T, Tipparaju SM, Moreno WA, Pathak Y (2016) Introduction to artificial neural network (ANN) as a predictive tool for drug design, discovery, delivery, and disposition: basic concepts, and modeling. In: Artificial neural network for drug design, delivery, and disposition. Academic Press, Cambridge, pp 3–13

Reinemann DJ & Helgren JM (2004) Online milk sensing issues for automatic milking. Paper number: 04–4191, ASAE/CSAE Annual International Meeting, Ottawa, Ontario, Canada

Salles MSV, da Silva SC, Salles FA, Roma LC Jr, El Faro L, Mac Lean PAB et al (2016) Mapping the body surface temperature of cattle by infrared thermography. J Therm Biol 62:63–69. https://doi.org/10.1016/j.jtherbio.2016.10.003

Schön PC, Puppe B, Manteuffel G (2001) Linear prediction coding analysis, and self-organizing feature map as tools to classify stress calls of domestic pigs (Sus scrofa). J Acoust Soc Am 110(3):1425–1431

Schön PC, Puppe B, Manteuffel G (2004) Automated recording of stress vocalisations as a tool to document impaired welfare in pigs. Anim Welf 13:105–110. https://doi.org/10.1017/S096272860002683X

Sellier N, Guettier E, Staub C (2014) A review of methods to measure animal body temperature in precision farming. Am J Agri Sci Technol 2:74–99

Sickel H, Ihse M, Norderhaug A, Sickel MA (2004) How to monitor semi-natural key habitats in relation to grazing preferences of cattle in mountain summer farming areas: an aerial photo, and GPS method study. Landsc Urban Plan 67(1–4):67–77

Suryawanshi KR, Redpath SM, Bhatnagar YV, Ramakrishnan U, Chaturvedi V, Smout SC, Mishra C (2017) Impact of wild prey availability on livestock predation by snow leopards. R Soc Open Sci 4(6):170026. https://doi.org/10.1098/rsos.170026

Tangorra FM, Calcante A (2018) Energy consumption, and technical-economic analysis of an automatic feeding system for dairy farms: results from a field test. J Agri Eng 49(869):228–232

Williams ML, Mac Parthaláin N, Brewer P, James WPJ, Rose MT (2016) A novel behavioral model of the pasture-based dairy cow from GPS data using data mining, and machine learning techniques. J Dairy Sci 99(3):2063–2075. https://doi.org/10.3168/jds.2015-10254

Yoshihara Y, Oya K (2021) Characterization, and assessment of vocalization responses of cows to different physiological states. J Appl Anim Res 49(1):347–351. https://doi.org/10.1080/09712119.2021.1967756

Chapter 22
Artificial Intelligence-Driven, IoT-Based Technologies in Agriculture: A Review

Bhawanpreet Kaur and Chandra Sekhar Mukhopadhyay

Abstract Artificial intelligence (AI), the most coveted cutting-edge technological intervention, empowers agricultural and animal farming sectors in many ways. The application of AI in farm sectors applies complex algorithms to use the Internet of things (IoT), mechatronics, robotics, information and communication technologies (ICTs), and intelligent devices to achieve the target of smart farming. IoT-oriented gadgets and AI-enabled devices are embedded with sensors to connect to and interact via the Internet. IoT also plays a crucial role in agriculture applications, making it possible for farmers and agripreneurs to collect valuable data. The present review provides an interdisciplinary overview of the latest development and contribution of AI-enabled gadgets and mechatronics in smart farming. AI has transformed the agriculture industry by complementing the current farming practices with smart services and automated decision-making. Such technological interventions are anticipated to result in improved agricultural practices and yields and qualitative improvement in farmers' lives.

Keywords Agriculture · Smart farming · Robots · Artificial intelligence · Internet of things

22.1 Introduction

Agriculture is the linchpin of the economy of any nation. India is one of the world's largest producers of several agricultural commodities. The set of traditional practices in agriculture is collectively called traditional agriculture practices (TAPs). Traditional agriculture has been practiced since prehistoric times and is approximately 4000 years old (Sofia et al. 2006). It is fundamental and essential in the stable growth and sustainable development of the economy of any nation (Johnston and Kilby

B. Kaur · C. S. Mukhopadhyay (✉)
Department of Bioinformatics, College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India
e-mail: csmukhopadhyay@gadvasu.in

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_22

1975). With the advent of computational technologies, robotics, and information and communication technologies (ICTs), traditional farming practices have increased swiftly. As the mechanized agricultural practices have been smartened up by artificial intelligence- (AI-) driven automated systems, it has cut down the risks of losses. It has enhanced sustainable development, crop productivity, and animal production. Sensors attached to animals and plants, data received through drones and satellites, and unmanned ground and aerial vehicles are utilized as input data for the data analyses. In general, machine learning (ML), an integral component of AI, provides a machine the ability to learn automatically and improve from knowledge and experience. It involves teaching a computer to recognize data patterns. ML, thereby, creates rules for analyzing the input data rather than traditional programming.

22.2 Big Data Analysis Using Artificial Intelligence

Big data analysis refers to the field of computational science that analyzes and extracts from "huge and complex" data, which are otherwise impossible to analyze using traditional data analysis methods. Big data has a significant impact on smart farming. It includes the conceptual framework in the farm processes and farm management, which contains basic management functions like sensing and monitoring, analysis and decision making, and intervention. Machine learning (ML) is a subfield of AI that confers a machine the ability to learn by using data. It follows the learning process of the human brain. ML helps to train a machine to improvise the analysis of complex data by finding hidden patterns and precise solutions (Tantug and Ve Türkmenoğlu 2015), and the application of ML is shown in Fig. 22.1.

Trained machines demonstrate AI. The researchers drew a parallel between the human brain and computing machines and hypothesized the working principle of the artificial neurones (http://www2.psych.utoronto.ca/users/reingold/courses/ai/cache/neural4.html).The term "Artificial Intelligence" was first coined by John McCarthy in 1955 (McCarthy 2020). The domain of "artificial intelligence" was born at the "Dartmouth Summer Research Project on Artificial Intelligence" (Dharmaraj and Vijayanand 2018). For the last seven decades, AI has traversed a long path through sunny and rainy days. It has evidenced the "First AI winter" (1974–1980) and the "Second AI winter" (1987–1993) due to curtailed or no funding for AI projects. Now, it has been well perceived that AI has a significant impact on the quality of life. Figure 22.2 shows the several branches of AI.

It is considered that an artificial neural network (ANN) lays the foundation of deep learning (DL), which is a sub-domain of machine learning. DL imitates the information transmission protocol of the human brain and forms layers that receive and transfers information (Brownlee 2019). DL is used for identifying objects, removing language barriers, data processing, etc. The Internet of Things (IoT) refers to the different kinds of physical devices around the world that are embedded with sensors and connected to the Internet. The IoT and AI are important technologies in today's world. When AI and IoT are combined, the outcome is AIoT or

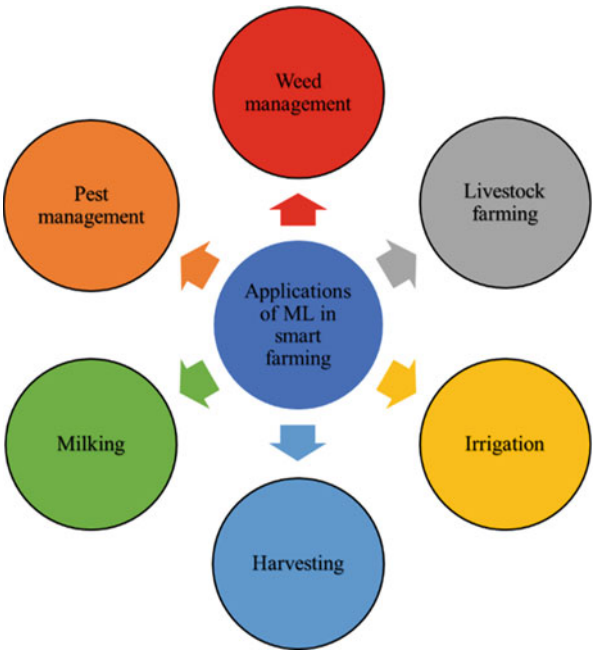

Fig. 22.1 Applications of machine learning (ML) in smart farming. AIoT: the portmanteau of AI and IoT

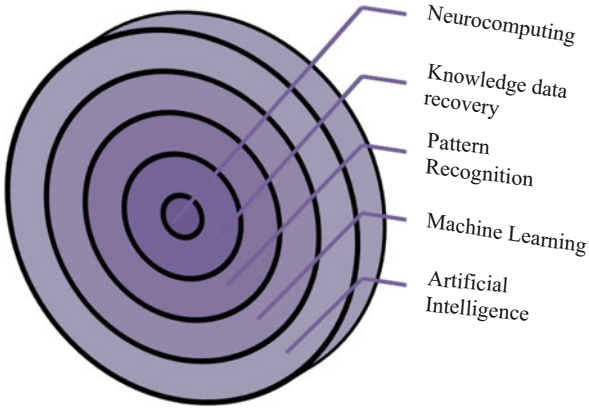

Fig. 22.2 Branching of AI

AI. Interestingly, the peripheral nervous system (of humans) is represented by the devices in the IoT, while AI is represented by the central nervous system (CNS).

22.3 Smart Farming or Digital Farming or iFarming

Smart farming is also called "intelligent farming." In the areas of agriculture and animals, it is the '*Fourth Industrial Revolution.*' It is an evolving model of farming which enables a farmer to use cutting-edge digital technologies of information &

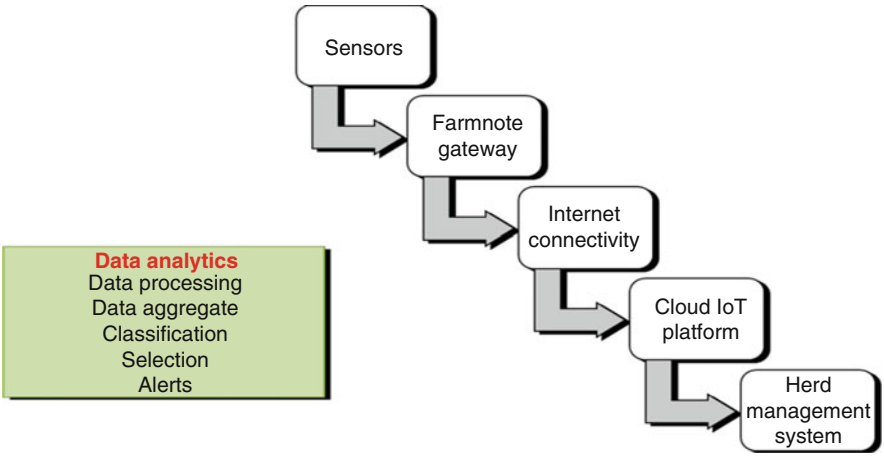

Fig. 22.3 Using IoT for the development of smart dairy farming

Table 22.1 Some of the most often used physical sensors used in smart farming

Sensor	Used	Applications	References
Parrot sequoia	Growth check	Crop assessment	Xue et al. (2019)
E-201	pH check	Substrate monitoring	Guandong et al. (2018)
TSL2561	Luminosity	Environment monitoring	Syafarinda et al. (2018)
Blueberry RFID reader	Tracking	Other	Ko et al. (2014)

communication technologies (ICT) to manage farms to maximize return and optimize the use of the human intervention (Walter et al. 2017). Smart farming uses IoT-based technologies in agricultural production to reduce waste and boost productivity. Many machine learning algorithms belonging to artificial neural networks are being used more efficiently to enable the machine to take decisions on its own, recognize and evaluate images and enhance the prediction accuracies with minimum follies. IoT is used for development of Smart Dairy Farming, as described in Fig. 22.3 (Table 22.1).

22.4 Agricultural Drones: Solution to Agricultural Problems

The major advantages of using drones are imaging of crop health, field irrigation, crop monitoring, spraying, planting, soil type, field assessment, plant health indices, plant counting and yield prediction, plant height measurement, nitrogen content, drainage, and weed mapping, and many more. The strategy and planning rely on

real-time data collection, processing, and drone technology will give high-tech applications in the agricultural sector.

Drones do supervision tasks with the help of gathering technologies like infrared technology (IFT), infrared cameras, a global positioning system (GPS), multispectral images, and laser (consumer, commercial and military UAV), which can cover vast acres of land in one move. Drones are controlled by ground control systems (GSC) and are also called "*ground cockpits*". Drones come in various sizes and types, including rotary drones fixed wing drones. Different types of cameras are used visual cameras, thermal cameras, multispectral cameras, Light detection, and ranging (LiDAR). For example, predator drones are used for military purposes.

Before the crop cycle, the drone can be used to produce a three-dimensional (3D) field map of detailed terrain, drainage, soil viability, and irrigation. Nitrogen-level management (NLM) can also be done by drone. In the case of cattle, the number of animals in a herd, their weight, movements, and possible anomalies such as lameness. Different sensors and software are used in the studies to detect crop diseases or classification, as shown in Table 22.2.

22.5 Two Major Categories of Agriculture

22.5.1 Precision Agriculture or Precision Farming

In precision farming, plants and cattle get precisely treated and determined by machines with strong accuracy. By precisely measuring variations within a field, farmers can boost the effectiveness of pesticides and fertilizers. Through the studies, it has been reported that in October 2016, the United States Department of Agriculture (USDA) announced the role of precision agriculture technologies and increased profits and net returns (Schimmelpfennig 2016).

22.5.2 Precision Livestock Farming

Precision livestock farming (PLF)

enables farmers to better monitor the needs of individual animals and to adjust their nutrition accordingly, thereby preventing disease and enhancing herd health. With this information, sick animals can be identified. To prevent the spread of disease, they can be separated from the herd. Large farm owners can use wireless IoT applications to monitor their cattle's location, well-being, and health. With this information, they can identify sick animals so that they can be separated from them to prevent the spread of disease (Fournel et al. 2017). Precision livestock farming includes cameras, accelerometers, gyroscopes, radio-frequency identification systems, pedometers, and optical and temperature sensors (Li et al. 2020). Several important components of precision farming are shown in Fig. 22.4.

Table 22.2 Advanced technologies used in smart farming

Name of technology	Features
Geographic information system (GIS)	• It analyses the present geographic data and comprises hardware and software supporting the compilation, storage, retrieval, and analysis of yield, soil survey, crop type, nutrient levels, and prone of geographical attributes and location data to produce maps • GIS maps contain layers of information like pests
Global positioning system (GPS)	• GPS is a network of orbiting satellites that obtain a bird-eye point of view • It enables farmers to identify field information and make decisions regarding seed plantation, herbicides, pesticides, fertilizers, and irrigational needs
Grid soil sampling and variable-rate fertilizer (VRT) application	• It is an unbiased, simple, and relatively quick method for site-specific soil management • Grid soil samples are analyzed and interpreted to determine crop nutrient needs
Rate controllers	• Controllers control the delivery rate of chemical inputs like fertilizers and pesticides • It monitors the speed of the tractor/sprayer as well as the flow rate across the field
Mobile apps or applications	• Using devices like smartphones, tablets, etc., and the availability of internet connectivity, it is very easy to share or get any information from anywhere. • Applications can also provide predictive weather forecasts, a variety of seedlings available, fertilizer, and many more
Sensors	• It is an integral part of precision agriculture—Location sensors, optical sensors, electrochemical sensors, etc.

The current studies were based on the criteria like crop monitoring, applications of IoT, and different sensors and software used in the studies to detect crop diseases or classification, as shown in Table 22.3 and the currently used agriculture techniques are shown in Table 22.4.

22.6 How Can AI Revolutionize Agriculture and Animal Sciences?

- **Outbreak prediction**: It helps to predict the possible outbreak of pandemics, like Coronavirus, Ebola, Monkey-pox, Lumpy skin disease (LSD), African swine fever (ASF), etc.
- **Medical/veterinary image diagnosis**: AI and IoT help imaging and data analysis in various techniques like USG, MRI images, and skin cancer prediction.

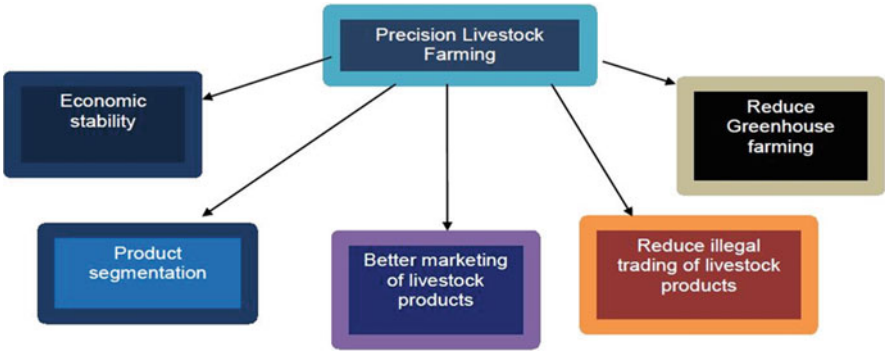

Fig. 22.4 Some of the essential components of precision livestock farming

Table 22.3 Recent research using IoT in crop management

Aim	Problem area	Tools used	Outcome	References
IoT-based smart crop monitoring	Using different sensors. Global system for mobile communications (GSM) technology used	Soil moisture sensor, temperature and humidity sensor	Outputs are checked with dry soil samples as well as wet soil sample	Balaji et al. (2019)
IoT solution for plant monitoring in smart agriculture	Precision agriculture. Power-efficient. Solar panel supplied IoT system for productive agricultural process	Sensors for temperature, leaf wetness, relative humidity of air and soil, and temperature of the soil	The output was better grape quality. With lower costs. Power efficient	Marcu et al. (2019)
Disease detection in tomato plants and remote monitoring of agricultural parameter	Remote monitoring of tomato plants. Detection of different types of soil and moisture content, disease classification	Using the internet of things (IoT), deep learning (DL), temperature and humidity sensor, and moisture sensor. NodeMCU, Blynk cloud	Live image tested classified correctly 42/50 images. (Diseased or healthy) 84% achieved accuracy, 95% training accuracy, 92% validation accuracy	Deepak et al. (2019)
Design of Novel Wireless Sensor Network Enabled IoT-based smart health monitoring system for thicket of trees	Low power consumption and used solar power to avoid power breaks. Data analysis based on cloud computing	Raspberry pi 3, deep learning (DL), universal software radio peripheral (USRP), GNU radio, sensors used	Data analytics and real-time surveillance, sensors analyzed at different environmental conditions	Sridhar et al. (2020)

Table 22.4 AI-enabled agriculture techniques being used in advanced nations

Used Technology	Advantages	Year	References
AI, sensors, GSM, Bayes algorithm	For seed sowing, proposed IoT and AI-based "AGROBOT". The monitors' weather forecast, fertilizers & pesticides need water requirements. Reduces labor costs and enhances crop yield	2020	Ragavi et al. (2020)
AI, IoT-based sensors, solar power, Arduino, raspberry pi, cloud computing	Cloud storage users can access data at any time	2020	Sinwar et al. (2020)
Solar energy system, image processing technique	Prevents data loss	2020	Gulec et al. (2020)
Wireless sensor networks (WSN), cyber-physical system (CPS), cloud computing, machine learning using K-means clustering	Helps in the management of water in agriculture. Web- and mobile-based applications, motion detection devices	2020	Saad et al. (Saad and Gamatié 2020)
IoT-AI based tools	E-agricultural solution. AI and IoT are used to record soil & temperature conditions. AI-based algorithms are used to prioritize crop location, soil type	2019	Rajesh et al. (2019)
SATURAS and stem water potential (SWP) technology	SATURAS and stem water potential (SWP) technology-based mechanism is introduced for water requirements	2019	Divya et al. (2019)
IoT, data mining, data analytics, big data storage and analytics, cloud computing	Platform and design help access to IoT, and improve crop productivity	2019	Farooq et al. (2019)
IoT, Wi-fi, ZigBee, sensors	An automated greenhouse control system with sensors for agricultural production. It is beneficial to small-scale farmers and improves the agricultural environment	2019	Gowri (2019)
Image processing using AI and IoT	For effective agricultural activities, a system based on IoT, AI, and image processing was proposed. Drones are used to collect data in real-time in the field	2019	Bhatta and Thangadurai (2019)
Wireless sensor network, cloud-based IoT architecture	The efficiency, productivity, and profitability of many agricultural production systems	2019	Anushree and Krishna (2018)
Real-time processing, cloud computing	Gives a faster response and less network cost, and reduced energy consumption	2018	Aher et al. (2018)
An analytical method, applied economics. Big data, precision agriculture	Crop disease, agricultural policy, climate forecasting, crop yield, and crop selection	2018	Coble et al. (2018)

(continued)

Table 22.4 (continued)

Used Technology	Advantages	Year	References
IoT, AI-based sensors, Arduino, and Bluetooth, are all about the internet of things	Proposed cost-effective automatic irrigation system (AIS) based on real-time soil moisture content	2019	Kumar et al. (2018)
Data aggregation at layer 1 (sensor network), layer 2 (base station – Internet), and layer 3 (response center server)	Reduces the complexity of data analysis at the top layer	2018	Kahtani and Karim (2018)
Big data analytics using global positioning system (GPS), yield monitoring and mapping, information management, variable-rate technology (VRT) technologies	Highly efficient and accurate technology	2017	Gill et al. (2017)
IoT, cloud computing, IoT-based smart farming and precision agriculture	The device is simple to operate and can be applied on a small as well as large scale	2017	Nuvvula et al. (2017)
Large data and metadata, sensor cloud infrastructure	Used in the field of farming	2016	Ravisankar et al. (2017)
Big data sensor, cloud infrastructure	Mobile decision system, efficient IoT-based tool	2016	Nandyala and Kim (2016)

- **Crops production:** Smart farming systems also enable careful management of the field and production of crops with qualitative and quantitative targets (Yvoz et al. 2020).
- **Weeds prediction**: The presence of unwanted grass in the crops is tracked by using a robotics system. ML technologies collaborate with imaging or non-imaging spectroscopy that can allow for differentiation and localization of target weeds, enabling precise application of herbicides to specific zones instead of spraying the entire file (Slaughter and Giles 2008).
- **Enhanced product quality**: The production process is controlled and higher standards of crop quality and growth capacity is maintained. The most common indices for harvesting include fruit quality, solids content, and skin color (Papageorgiou et al. 2018).
- **Health monitoring of crops**: 3D laser scanning is essential to building crop metrics. The technology will also be used to monitor crops.
- **Soil management**: Soil is a natural resource for owning better control like land degradation, imbalance of soil-nutrient, fertilizers overuse, soil erosion, irregular crop rotations, livestock overgrazing, etc. (Chasek et al. 2015).
- **Fertilizer & pesticide management:** Pesticide levels on crops can help farmers better yield. Farmers can understand pests and their activity, location, and patterns in management practices.
- **Irrigation system management**: Automated irrigation system understands the water level needs of a plant. It can reduce human interference, and decisions are made whether fields to irrigate or not.

- **Growth drive by IoT**: Proximity sensing and remote sensing are two technologies that are primarily used to improve growth using IoT. It also identifies relationships and mandates comprehensive datasets and powerful algorithms using Machine Learning (ML) (Khaki and Wang 2019).
- **AI in automated milking**: AI increases the application in animal husbandry. Intelligent sensors and automatic milking units help analyze milk quality and quantity (Kumari and Dhawal 2021).
- **Animal disease prediction:** To prevent disease outbreaks, even before a large-scale outbreak, AI and machine learning algorithms help the farmer to predict the change in the cow's behavior, deviations, or abnormalities.
- **Detection of oestrus cycle:** AI also provided insights into the heat stress, change in feeding efficiency, and the oestrus cycle.
- **Improving animal health**: Using facial recognition, various practical applications help to understand the animal's emotional and attentional state. As further advantages, we may find injuries, diseases, or signals of predator attacks.
- **Livestock monitoring:** IoT devices can be used to identify sick animals, thereby preventing the spread of diseases.

22.7 Scope of AI in Agriculture

Agriculture has been tackling more important difficulties like the absence of irrigation systems, temperature change, food scarcity, etc. At regular intervals, AI-based technology will be helpful in helping farmers in high yielding and better seasonal crops. In many countries, including India, farmers depend on the monsoon for their cultivation. AI and IoT have the potential to record information about reproduction, health, and nutrition. The sensor also provides farmers with solutions for each individual cow.

22.8 Conclusion and Future Prospectives

The recent, advanced technologies, including neural networking, IoT, digital image processing (DPI), and the concept of logic gates (LG), are vital in intelligent farming. The condition of the environment will help to understand the current situation of the agricultural lands. Thus, in this way, the agricultural products in both qualitative and quantitative methods are increased. This will help the entire world with the necessary food items in an efficient manner. This will also enhance the competitiveness and sustainability of their productions. With the population growing rapidly, the demand can be successfully met if the ranchers, as well as small farmers, implement agricultural IoT solutions in a prosperous manner.

Acknowledgments The authors thankfully acknowledge the funding provided by the Department of Biotechnology, Government of India, through the collaborative research project "Parentage Determination and Cytogenetic Profiling in Dogs (DBT-19I)".

Competing Interests

The authors declare that they have no competing interests.

Declaration Bhawanpreet Kaur and CS Mukhopadhyay designed the outline of the review article. Bhawanpreet and CS Mukhopadhyay wrote the manuscript. Manuscript editing and proofreading of the manuscript was done by CS Mukhopadhyay. All authors read and approved the manuscript.

Ethics Approval and Consent to Participate Not applicable.

References

Aher A, Kasar J, Ahuja P, Jadhav V (2018) Smart agriculture using clustering and IOT. Int Res J Eng Technol 5(03):4065–4068

Anushree MK, Krishna R (2018) A smart farming system using Arduino based technology. Int J Adv Res Ideas Innov Technol 4(4):850–856

Balaji GN, Nandhini V, Mithra S, Priya N, Naveena R (2019) IoT-based smart crop monitoring in farmland. Imp J Interdiscip Res 4:88–92

Bhatta NP, Thangadurai N (2019) Utilization of IoT and AI for agriculture applications. IJEAT 5: 2731–2735

Jason Brownlee (2019) What is deep learning? https://machinelearningmastery.com/what-is-deep-learning/. Accessed 14 Dec 2020. https://doi.org/10.3390/nu11112615

Chasek P, Safriel U, Shikongo S, Fuhrman VF (2015) Operationalizing zero net land degradation: the next stage in international efforts to combat desertification? J Arid Environ 112:5–13. https://doi.org/10.1016/j.jaridenv.2014.05.020

Dharmaraj V, Vijayanand C (2018) Artificial Intelligence (AI) in agriculture. Int J Curr Microbiol App Sci 7(12):2122. https://doi.org/10.20546/ijcmas.2018.712.241

Divya R, Chinnaiyan R (2019) Reliable AI-based smart sensors for managing irrigation resources in agriculture—a review. In: International Conference on Computer Networks and Communication Technologies: ICCNCT 2018. Springer, Singapore, pp 263–274

Farooq MS, Riaz S, Abid A, Abid K, Naeem MA (2019) A survey on the role of IoT in agriculture for the implementation of smart farming. IEEE Access 7(1):156237–156271

Fournel S, Rousseau AN, Laberge B (2017) Rethinking environment control strategy of confined animal housing systems through precision livestock farming. Biosyst Eng 155:96–123. https://doi.org/10.1016/j.biosystemseng.2016.12.005

Gill S, Chana I, Buyya R (2017) IoT based agriculture as a cloud and big data service: the beginning of digital India. J Organ End User Comput 29(4):1–23. https://doi.org/10.4018/JOEUC.2017100101

Gowri KR (2019) Greenhouse monitoring and scheming based IoT technology. EPRA Int J Res Dev 4(4):316–321

Guandong G, Yuchen J, Ke X (2018) An IOT-based multi-sensor ecological shared farmland management system. Int J Online Eng 14:81. https://doi.org/10.3991/ijoe.v14i03.8199

Gulec O, Haytaoglu E, Tokat S (2020) A novel distributed CDS algorithm for extending lifetime of WSNs with solar energy harvester nodes for smart agriculture applications. IEEE Access 8(12): 58859–58873

Johnston B, Kilby P (1975) Agriculture and structural transformation: economic strategies in late developing countries. Oxford University Press, oxford

Kahtani AI, Karim L (2018) Dynamic data aggregation approach for sensor-based big data. Int J Adv Comput Sci Appl 9(7):62–72

Khaki S, Wang L (2019) Crop yield prediction using deep neural networks. Front Plant Sci 10:621

Ko D, Kwak Y, Song S (2014) Real time traceability and monitoring system for agricultural products based on wireless sensor network. Int J Distrib Sens Netw 10:832510. https://doi.org/10.1155/2014/832510

Kumar S, Raja P, Bhargavi G (2018) A comparative study on modern smart irrigation system and monitoring the field by using IOT. In: IEEE

Kumari M, Dhawal K (2021) Application of Artificial Intelligence (AI) in animal husbandry. Vigyan Varta 2(2):27–29

Li N, Ren Z, Li D, Zeng L (2020) Review: automated techniques for monitoring the behaviour and welfare of broilers and laying hens: towards the goal of precision livestock farming. Animal 14: 617–625. https://doi.org/10.1017/S1751731119002155

Marcu I, Suciu G, Bălăceanu C, Drăgulinescu A and Dobrea MA (2019) IOT Solution for Plant Monitoring in Smart Agriculture. In: 2019 IEEE 25th international symposium for design and Technology in Electronic Packaging (SIITME), Cluj-Napoca, Romania, pp 194–197. doi: https://doi.org/10.1109/SIITME47687.2019.8990798

John McCarthy (2020) Computer scientist known as the father of AI. The Independent. https://www.independent.co.uk/news/obituaries/john-mccarthy computer-scientist-known-father-ai-6255307.html. Accessed 14 Dec 2020

Nandyala CS, Kim HK (2016) Big and meta data management for U-agriculture mobile sevices. J Software Eng App 10:257

Nuvvula J, Adiraju S, Mubin S, Shahana B, Valisetty V (2017) Environmental smart agriculture monitoring system using internet of things. Int J Pure Appl Math 115(6):313–320

Papageorgiou EI, Aggelopoulou K, Gemtos TA, Nanos GD (2018) Development and evaluation of a fuzzy inference system and a neuro-fuzzy inference system for grading apple quality. Appl Artif Intell 32:253–280. https://doi.org/10.1080/08839514.2018.1448072

Ragavi B, Pavithra L, Sandhiyadevi P, Mohanapriya GK, Harikirubha S (2020) Smart agriculture with AI sensor by using Agrobot. In: IEEE

Rajesh T, Kumar KT, Balan MM, Sidhartha MS (2019) E-agriculture based on AI and IoT. Int J Res Appl Sci Eng Technol 7:343–347. https://doi.org/10.22214/ijraset.2019.2038

Ravisankar K, Sidhardha K, Prabadevi B (2017) Analysis of agricultural data using big data analytics. J Chem Pharm Sci 10(3):1132–1135

Saad A, Gamatié A (2020) Water Management in Agriculture: a survey on current challenges and technological solutions. IEEE Access 8(1):38082–38097

Schimmelpfennig D (2016) Farm profits and adoption of precision agriculture. USDA, 217, 1–46. Services. Int J Software Eng Appl 10(1):257–270

Sinwar D, Dhaka VS, Sharma MK, Rani G (2020) "AI-based yield prediction and smart irrigation," in the internet of things and analytics for agriculture, vol 2. Springer, Singapore, pp 155–180

Slaughter DC, Giles DK, Downey D (2008) Autonomous robotic weed control systems: a review. Comput Electron Agric 61:63–78. https://doi.org/10.1016/j.compag.2007.05.008

Sofia PK, Prasad R, Vijay VK (2006) Organic farming-tradition reinvented. Indian J Tradit Knowl 5(1):139–142

Sridhar B, Sridhar S and Nanchariah V (2020) Design of novel wireless sensor network enabled IOT based smart health monitoring system for thicket of trees. 2020 fourth international conference on computing methodologies and communication (ICCMC), Erode, India, pp 872–875. https://doi.org/10.1109/ICCMC48092.2020.ICCMC-000161

Syafarinda Y, Akhadin F, Fitri ZE, Widiawan B, Rosdiana E (2018) The precision agriculture based on wireless sensor network with MQTT protocol. IOP Conf Ser Earth Environ Sci 207:012059. https://doi.org/10.1088/1755-1315/207/1/012059

Tantug AC, Ve Türkmenoğlu C (2015) Türkçe Metinlerde Duygu Analizi. Yüksek lisans tezi, İstanbul Teknik Üniversitesi, İstanbul

Walter A, Finger R, Huber R, Buchmann N (2017) Opinion: smart farming is key to developing sustainable agriculture. PNAS 114(24):6148–6150

Xue J, Fan Y, Su B, Fuentes S (2019) Assessment of canopy vigor information from kiwifruit plants based on a digital surface model from unmanned aerial vehicle imagery. Int J Agric Biol Eng 12: 165–171. https://doi.org/10.25165/j.ijabe.20191201.4634

Yvoz S, Petit S, Biju-Duval L, Cordeau S (2020) A framework to type crop management strategies within a production situation to improve the comprehension of weed communities. Eur J Agron 115:126009. https://doi.org/10.1016/j.eja.2020.126009

Chapter 23
Revolutionize One Health Through Quantum Computing

Ritwika Das and Dwijesh Chandra Mishra

Abstract Quantum computing is an emerging and ongoing research area dependent on both quantum mechanics and computer science. With the generation of a large amount of data in different areas, conventional computing algorithms will fail to analyze them efficiently in a short period. However, computation using qubits and quantum properties like superposition, entanglement, and parallelism has the potential to overcome these problems by providing robust and faster solutions. Quantum computing is highly efficient to revolutionize medical science in the coming days. Given this, basic features and potential applications of quantum computing in healthcare and medical science are discussed in this article.

Keywords Quantum computing · One health · Health-care · Classical computing

23.1 Introduction

The invention of the computer is considered to be the greatest, life-changing innovation in the history of mankind and has become an indispensable part of our daily life. Rapid advancement in the computing system has assisted in solving complex and diverse problems in every domain of science, especially, life sciences. From genome sequencing to predicting the behaviour of several biomolecules, computing algorithms are being routinely used to extract and analyze information from biological experiments. Despite various advantages of a classical computing system, these methods suffer from several limitations like size constraint, energy

R. Das
Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, PUSA, New Delhi, India
e-mail: Ritwika.Das@icar.gov.in

D. C. Mishra (✉)
Division of Agricultural Bioinformatics, ICAR - Indian Agricultural Statistics Research Institute (IASRI), Library Avenue, PUSA, New Delhi, India
e-mail: Dwijesh.Mishra@icar.gov.in

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_23

consumption, lack of improvement opportunities, inability to solve large and complex problems, etc. To solve these shortcomings, a new technique, quantum computing has been introduced. In the 1980s, Richard Feynman and Yuri Manin independently proposed this new computing technique to develop new and highly powerful computers which are supposed to perform operations that are not possible in classical computers in a fundamentally different way. Quantum computing has emerged based on theories of quantum mechanics. It has features like superposition and entanglement which are comparable with the angular momentum of electrons referring to the motion of electrons around the nucleus of an atom. Quantum computation in terms of qubits can exponentially increase the data processing speed and may solve complex problems. Challenges in building quantum computing-based hardware, i.e., quantum computers have not stopped the development of quantum-based optimization algorithms useful for molecular simulations, larger calculations, optimizing protein folding, image processing, drug discovery, and many others. Quantum computing has enormous potential to revolutionize the medical science and healthcare industries. In the following sections, the basic features and needs of quantum computing as well as some of their applications in healthcare are discussed in brief.

23.2 Why Quantum Biology?

Living entities are made up of atoms and molecules, and fundamentally, molecular mechanisms are described in the light of quantum mechanics. In most cases, subtle quantum effects do not play a significant role in defining overall biological function. A biological system is mostly explained using theories of classical mechanics and the differences between classical approximations and quantum-mechanical theories are generally neglected. We know that biological processes are complex and highly dynamic. The dynamics in biological systems such as chemical reactions, light absorption, formation of excited electronic states, transfer of excitation energy, etc. are mediated by electrons and protons at the atomic level. So, we can say that all biological processes are quantum mechanical. While quantum effects are difficult to observe on macroscopic time and length scales, quantum mechanics is necessary for better and more precise characterization of the behaviour of biological processes and subsystems on nanoscales. Hence, quantum biology is an emerging scientific field indicating the applications of quantum mechanics and theoretical chemistry to biological objects and problems. Biological processes like photosynthesis, magnetoreception, olfaction, enzyme catalysis, respiration, and neurotransmission can be well understood using quantum biology.

23.3 Classical Computing and its Limitations

The advancement of low-cost next-generation sequencing technologies in the past decade has generated an enormous amount of sequence data to facilitate the understanding of biological processes at the macromolecular level, i.e., DNA, RNA, and protein. In the 2000s, the concept of big data has been highly popular because of the complexity, heterogeneity, and rapid growth of biological data. Massive growth in the technologies and data science has helped to analyze these data using highly powerful computers in minutes aiding research in areas of bioinformatics, computational biology, artificial intelligence, drug designing, medical science, etc. In a conventional computer, these computations are performed in transistors, i.e., silicon-made semiconductor chips, and the fundamental unit of information is represented by Boolean bits, i.e., either 0 or 1. For example, if the current is passing through the chip, then it is said to be in a logical 1 (on) state or logical 0 (off) state. Improvement in classical computing has facilitated the development of highly powerful and miniaturized chips which made classical computers more compact and faster. According to Intel co-founder Moore (1965), the computation capacity of semiconductor chips gets doubled after every 18 months. However, when the chip size is very small, viz., ~10 nm, electrons start to reveal quantum nature which cannot be explained by the principles of classical physics (Barde et al. 2011). Hence, the computing power of these computers is not infinite. Some significant limitations of classical computing are listed below:

1. **Energy consumption**: Transistors function by switching between the on (1) and off (0) states mediated by hundreds of electrons. A large amount of heat is generated during this process. The speed of switching between states depends upon the size of the transistor.
2. **Size constraint**: When the size of a transistor is very small (<1 nm), the flow of electrons in the off state gets disturbed leading to the process of quantum tunneling (Seabaugh 2013).
3. **Little room for further improvement**: To make conventional computers faster, the sizes of semiconductor chips are minimized to reduce the time of transfer of instructions among chips by minimizing their internal distances. The process demands to fit a higher number of transistors on a chip making it denser than before. Furthermore, shrinking in chip size demands for more complex designs to be made. But it has been predicted that shortly the increase in transistors densities on shrinking size of chips will come to an end (Hruska 2013).
4. **Inability to explain underlying quantum nature of biological processes**: Many biological processes like light harvesting during photosynthesis, bird navigation through magnetoreception, olfactory reception, etc. can be well explained by theories of quantum biology like superposition, quantum coherence, quantum tunneling, etc. Conventional computer algorithms are unable to explain this underlying quantum mechanism as principles of classical computing are solely based on classical mechanics' inability to explain the quantum nature of particles.

5. **Inability to perform highly complex computation within a speculated time**: In classical computers, data processing is performed in terms of two distinct bits 0 and 1, and processing power increases almost linearly with the increase in several semiconductor chips. Although supercomputers have high processing speed, they may take a lot of time to perform complex and large data analysis. Some highly complex computations like factorization of a 500-digit number, sorting a large database, etc. are even out of their computation capacity.

Scientists felt the need to mitigate such limitations of classical computers and develop new technologies to either bypass or embrace this quantum nature. Hence, a new era of digital computation, i.e., quantum computing has been introduced to apply the principles of quantum mechanics leading to the development of quantum computers.

23.4 Quantum Computing

Quantum computing is an interdisciplinary area of research including principles of quantum mechanics and computer science. It uses quantum properties of atoms such as superposition, entanglement, and interference to solve highly complex problems in a very short period. The term "Quantum Computing" has been coined by Richard Feynman in 1981. It has been speculated that in the future, a single transistor or semiconductor chip will be controlled by a single electron (single electron transistor; SIT) (Barde et al. 2011). This mechanism can only be explained using quantum theory. Big companies like IBM, Google, Microsoft, etc., are giving more and more emphasis on research and development in the area of quantum computing. In 2019, IBM developed the first commercial quantum computer, IBM Q System One, that combines both quantum and classical computations in a single platform to be used in various research and business applications like quantum science, cryogenic engineering, systems engineering, and industrial design (Lardinois 2019; Anonymous 2021).

23.4.1 Basic Features of Quantum Computing

The salient features of quantum computing have been delineated below:

23.4.1.1 Qubits

In quantum computing, each unit of information governed by two electronic states of an atom is represented by quantum bits or qubits. Unlike bits in classical computing, qubits can take 0, 1, or, any number of values in between them at the same time. It

represents a complex two-state system where the state space may contain an infinite number of possible states. Examples of such systems are the polarization of photons, electron spin (1/2 or − 1/2), excited states, ground states of an atom, etc.

23.4.1.2 Superposition

The ground state and excited state of an atom are represented as $|0\rangle$ and $|1\rangle$ according to Paul Dirac's Bra-ket notation. Similar to the laws of quantum mechanics, each qubit ($|\psi\rangle$) can take any values between 0 and 1 which can be represented as a superposition or linear combination of two basic states as follows:

$$|\psi\rangle = \alpha|0\rangle + \beta \mid 1\rangle$$

where α and β are two complex vectors and $|\alpha|^2 + |\beta|^2 = 1$. It indicates that electrons can exist at every location at the same time point with varying probabilities.

In classical computing, 2 bits of information can be represented as 00, 01, 10, and 11. Whereas, in quantum computing, 2 qubits can take any number of values resulting in an exponential increase in computing capacity. This feature helps in performing complex calculations faster than traditional methods.

23.4.1.3 Parallelism

The superposition feature has enabled a quantum processor to perform multiple complex calculations simultaneously. This process is referred to as quantum parallelism which provides enormous computational power as compared to classical computers. Factorization of a 500-digit number in a classical supercomputer may take billions of years, whereas; it can be done within only 1 year using a quantum computer. Similarly, sorting and searching large databases can also be done very fast using this parallelism feature of quantum computation.

23.4.1.4 Entanglement

According to Albert Einstein's "Spooky action at a distance", an entangled pair of electrons always spin in opposite directions and influence each other through time and space even when they are not physically connected. That means, without any external forces, the direction of spin of an atom can be mathematically calculated from the rest of the atoms entangled in space. Similar to this concept, the change of state of a qubit in quantum computing is dependent on the change of states of other qubits even when they are not spatially separated. The usefulness of this feature can be observed in secured teleportation, quantum cryptography, and others.

Superposition, parallelism, and entanglement features have enabled quantum computing to solve problems that classical computers are unable to solve. Even

Table 23.1 Comparison between classical computing and quantum computing (Source: Rasool et al. 2021)

Criteria	Classical computing	Quantum computing
Unit of computation	Bits; two distinct states 0 and 1	Qubits; infinitely many states between 0 and 1 simultaneously
Computation power	Increases linearly with the number of chips	Has the potential to increase exponentially with an increase in qubits
Error rate	Low	High
Working environment	Can be operated at room temperature	An ultracold environment is needed for its operation
Applicability	Routine processing; not suitable for computer-intensive tasks	Suitable to perform computer-intensive complex large data analysis

though it is not being applied on a large scale till now, some service providers have started quantum computation in cloud computing platforms, viz., Amazon Bracket (Anonymous 2022). Recently, a report has been published (Porter 2019) which states that Google's qubit Sycamore quantum processor has successfully performed a calculation within 200 s which is possible to be done in around 10,000 years using a classical computer. The strengths, weaknesses, and applicability of both classical and quantum computing are represented in Table 23.1.

23.5 Applications of Quantum Computing

Quantum computing has enormous potential to bring out a revolution in various key sectors such as cryptography, financial modeling, precise weather forecasting, transportation, etc. Advanced quantum algorithms, as well as quantum computers, are being developed rapidly. Quantum computing can highly influence the healthcare sector. Massive computational capacity is highly useful for supersonic drug design, in silico clinical trials of drugs using simulated virtual human beings, high-speed DNA sequencing, predicting chronic diseases, securing medical data, developing new therapies and medicines, etc. (Fig. 23.1).

A quantum computer can create an efficient imaging system that will provide more clarity in disease diagnosis in real-time, and it can also perform complex optimization to figure out the optimum radiation plan for a cancer patient. Thus, quantum computation may improve the healthcare sector in terms of better healthcare plans, accelerated diagnoses, personalized medicine, and price optimization. Various application areas of quantum computing are discussed in detail in the following sections.

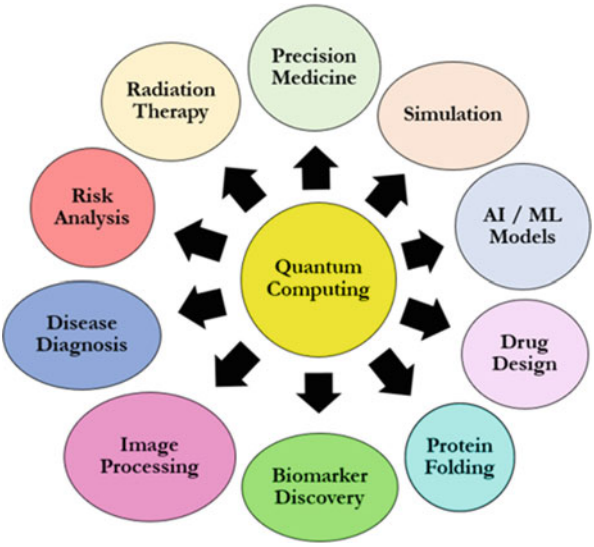

Fig. 23.1 Application of quantum computing in medical science and healthcare

23.5.1 Simulation and Optimization

Nowadays, medical science is highly dependent on computer algorithms based on machine learning, artificial intelligence, molecular simulations, etc. Quantum computing may improve the various classical binary machine learning methods for better analysis and prediction of big data and information processing. Complex correlations and dependencies among highly connected elements involving interactions between many electrons can be efficiently modeled by quantum computing algorithms. Machine learning models with quantum computational capability may solve complex big data problems efficiently and produce real-time results. Application of D-Wave quantum annealer for training and optimizing machine learning algorithms in hybrid classical-quantum computing system has several advantages like reduced training time, the small size of the training dataset, dimension reduction, etc. (Nath et al. 2021). Researchers have used D-Wave quantum annealing to solve various real-world problems (Fig. 23.2).

Higher resource requirements in complex simulation problems increase the time complexity of a problem exponentially which is easily manageable using quantum computing. In a recent study, a CNN-based hybrid model MERA (multi-scale entanglement renormalization ansatz) (Oh et al. 2020) has been proposed where a quantum convolution layer has been added to the CNN model replacing the original convolution layer and it is supposed to solve classification problem in quantum physics and chemistry with increased accuracy and precision.

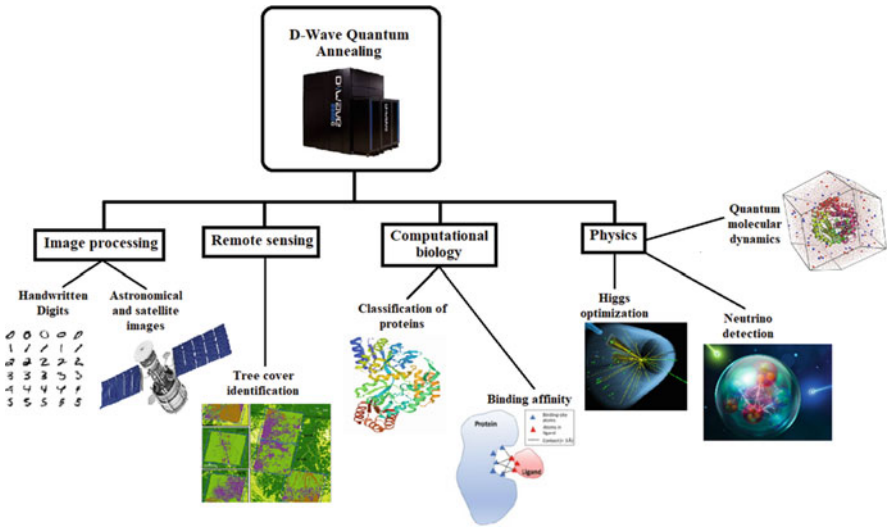

Fig. 23.2 Application areas of D-Wave quantum annealing

23.5.2 Image Processing and Disease Diagnosis

It has been observed that 10–20% of the total cost rises due to healthcare and maintenance, whereas the rest 80–90% occurs due to socio-economic factors, environmental aspects, and health-related behaviours (Rasool et al. 2021). Accurate diagnosis of a disease at an earlier stage leads to better treatment and may reduce the healthcare cost. In the past few years, computer-aided image diagnosis tools like X-ray, CT scans, and MRI are evolving at a faster rate and can detect disease earlier with high accuracy. These imaging techniques are highly useful in areas of neurobiology (Swan et al. 2022) and cancer research. However, these techniques have some limitations like experimental error, data quality, variability, replicability issues, etc. Quantum computing has the potential to overcome such limitations and provide better analysis of medical images such as edge detection. Quantum algorithms are being used in reconstructing medical images from MRI, CT scanners, and PET scanners (Kiani et al. 2020). Quantum machine learning has been applied to classify EEG data for Parkinson's disease patients to study Deep Brain Stimulation where 794 features have been analyzed from each of 21 EEG channels (Koch et al. 2019). QRNN (quantum recurrent neural network) with advanced processing techniques (Kullback–Leibler spatial patterns and Bayesian learning) has been found to outperform traditional Kalman filtering methods while tested upon real-time EEG data and BCI competition test data (Gandhi et al. 2014). In a study, it is found that a qutrit (three-level quantum states) model may analyze grayscale imaging data better than a qubit model using a quantum neural network model to segment brain lesions because many quantum states are not binary (Konar et al. 2021). Quantum-based

MRI tumor segmentation methods may use a quantum entropy classification method (Hasan et al. 2020) along with a quantum filtering technique (for noise reduction pre-processing) and a quantum artificial immune system-inspired SoftMax function in a deep spiking neural network (SNN) architecture (Ahmadi et al. 2021). Quantum algorithms can also be applied to analyze CT scans, e.g., classification of quantum data comparing COVID-19 and non-COVID-19 patient influenza and virus pneumonia lung CT scans, analyzed with TensorFlow Quantum and a D-Wave Systems quantum annealer (Sengupta and Srivastava 2021). Currently, single-cell methods are being widely used for disease diagnosis where advanced algorithms are required for the analysis of flow cytometry and single-cell sequencing data. Quantum-based machine learning algorithms like quantum supported SVMs may perform such analysis and boost the diagnosis method (Rasool et al. 2021).

23.5.3 Protein Folding and Drug Designing

The expression of a gene is reflected in terms of functions of the protein encoded from it. The function of a protein is expressed in its tertiary structural form, i.e., fold. Wrong folding is a possible cause of protein malfunction. Thorough research on the mechanism of protein folding may help in the new drug molecules and advanced therapies. Quantum computing may play a significant role in testing a large number of possible protein fold structures and recognizing the best fold simultaneously in a very short period. Again, researchers will also be able to model complex protein-protein interactions at the molecular level which is essential for better diagnosis of disease, its treatment, and subsequently new drug discovery. In drug designing, a potential drug molecule can be obtained when the underlying protein is having the conformation of the lowest energy. An extensive search algorithm is required to identify the suitable conformation out of zillions of possible conformations of that protein (Marx 2021). Peptide designing involves optimization of rotamers, i.e., side chains of proteins by searching and assessing various conformations of each rotamer. Packer module of Rosetta software does this rotamer optimization using the Monte Carlo simulation method. Scientists have built a D-Wave quantum enabled Packer module, QPacker where energies of individual rotamers and pairs of potential interacting rotamers are calculated using a classical method and quantum computing performs the searching through the whole set of all possible combinations of rotamers and finding out the best one (Marx 2021). It has been reported that using quantum technology, approximately 20,000 proteins can be encoded in the human genome and their interactions with existing drugs can be simulated (Rasool et al. 2021).

23.5.4 Radiotherapy

Radiation therapy is used for cancer treatment to kill cancerous cells and stop their multiplication. Highly precise computations are required in this process so that the radiation can affect only cancer-affected cells without causing any harm on healthy cells. Multiple complex calculations and simulations are needed to be performed to obtain an optimal radiation therapy. Quantum computing may speed up this process by performing multiple complex simulations simultaneously.

23.5.5 Precision Medicine

Precision medicine is the new era of medical science where prevention and treatment measures of a particular disease can be personalized. Here, the identification of relationships among causes and treatments of diseases is very important to predict the next course of action. Extensive research is going on in this regard. Presently, machine learning models can predict the risk associated with a future disease using data from electronic health records but these algorithms have some limitations like error, data quality, large feature space, the complex correlation between features, etc. Quantum-based machine learning models can be able to overcome these problems resulting in the early diagnosis of a disease. These algorithms may process continuously growing data with information on an individual basis, thus optimizing and personalizing healthcare services (Fig. 23.3).

23.5.6 Risk Analysis

Quantum computing helps in risk analysis by analyzing the current health of patients and predicting the tendency of the patient to be impacted by a particular disease. Based on this, insurance premiums can be optimized. The analysis of disease risks at the population level and intermixing them with the quantum risk models could help

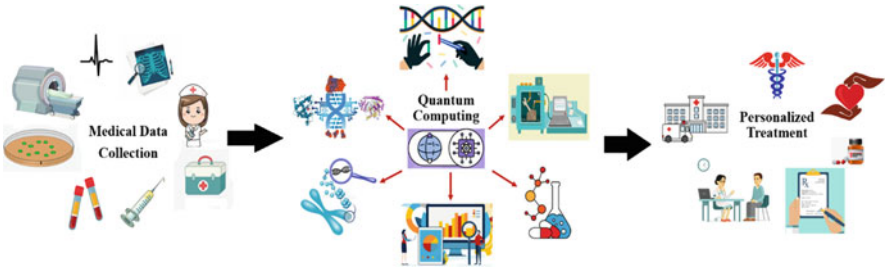

Fig. 23.3 Application of quantum computing in precision medicine

in computing financial risks and pricing models at a larger aspect. Quantum computing can perform classification and pattern recognition with high accuracy to detect malicious medical claims. Thus, it can efficiently help in the reduction of frauds in the healthcare sector and thereby saving billions of dollars of revenue (Rasool et al. 2021).

23.6 Conclusion

Quantum computing is a major step ahead compared to existing classical computing methods. Its unique features, i.e., superposition parallelism and enhancement have enabled us to solve large complex problems in just a couple of minutes. Various sectors of medical sciences like neurology, cardiology, vaccine research, etc. can be highly benefitted from quantum computing resulting in accelerated disease diagnosis, personalized medicine, drug discovery, automation, and many others. It is already being used in image processing such as CT scans, MRI, etc. Although there are so many advantages of quantum computing, there are some limitations too. Quantum algorithms and quantum computers are extremely difficult to build and program. Therefore, they are heavily affected by experimental errors, noisy outputs, and loss of quantum coherence crucial to their operation. Researches are still ongoing to overcome these shortcomings. In the future, more studies are needed to develop quantum-based advanced algorithms for big data analysis, ML/AI applications, large-scale optimization, quantum web and cloud services, simulation, quantum game theories, etc. Although quantum computing is still in its early stages, it leads us toward a new era of high-performance computing that will solve complex problems in biological data analysis like genome assembly, protein-DNA binding interactions, disease detection, and many more.

References

Ahmadi M, Sharifi A, Hassantabar S, Enayati S (2021) QAIS-DSNN: tumor area segmentation of MRI image with optimized quantum matched-filter technique and deep spiking neural network. Biomed Res Int 2021:1–16. https://doi.org/10.1155/2021/6653879

Anonymous (2021) IBM Quantum system one. https://research.ibm.com/interactive/system-one/

Anonymous (2022) Amazon Braket: accelerate quantum computing research. https://aws.amazon.com/braket/

Barde N, Thakur D, Bardapurkar P, Dalvi S (2011) Consequences and limitations of conventional computers and their solutions through quantum computers. Leonardo J Pract Technol 10(19): 161–171

Gandhi V, Prasad G, Coyle D, Behera L, McGinnity TM (2014) Quantum neural network-based EEG filtering for a brain-computer Interface. IEEE Trans Neural Netw Learn Syst 25(2): 278–288. https://doi.org/10.1109/TNNLS.2013.2274436

Hasan AM, Jalab HA, Ibrahim RW, Meziane F, Al-Shamasneh AR, Obaiys SJ (2020) MRI brain classification using the quantum entropy LBP and deep-learning-based features. Entropy 22(9): 1033. https://doi.org/10.3390/e22091033

Hruska J (2013) Intel's former chief architect: Moore's law will be dead within a decade. https://www.extremetech.com/computing/165331-intels-former-chief-architect-moores-law-will-be-dead-within-a-decade

Kiani BT, Villanyi A and Lloyd S (2020) Quantum medical imaging algorithms. https://doi.org/10.48550/arXiv.2004.02036. https://arxiv.org/abs/2004.02036

Koch M, Geraedts V, Wang H, Tannemaat M and Balck T (2019) Automated machine learning for EEG-based classification of Parkinson's disease patients. In proceedings of the 2019 IEEE international conference on big data (big data), Los Angeles, CA, USA, 9–12 December 2019

Konar D, Bhattacharyya S, Panigrahi BK, Behrman EC (2021) Qutrit-inspired fully self-supervised shallow quantum learning network for brain tumor segmentation. IEEE Trans Neural Netw Learn Syst 33:6331. https://doi.org/10.1109/TNNLS.2021.3077188

Lardinois F (2019) IBM unveils its first commercial quantum computer. https://techcrunch.com/2019/01/08/ibm-unveils-its-first-commercial-quantum-computer/

Marx V (2021) Biology begins to tangle with quantum computing. Nat Methods 18:715–719. https://doi.org/10.1038/s41592-021-01199-z

Moore GE (1965) Cramming more components onto integrated circuits. Electronics 38(8):114–117

Nath RK, Thapliyal H, Humble TS (2021) A review of machine learning classification using quantum annealing for real-world applications. Sn Comput Sci 2:1–11. https://doi.org/10.48550/arXiv210602964

Oh S, Choi J and Kim J (2020) A tutorial on quantum convolutional neural networks (QCNN). International Conference on Information and Communication Technology Convergence (ICTC), 2020. pp 236–239. https://doi.org/10.1109/ICTC49870.2020.9289439

Porter J. (2019) Google confirms 'quantum supremacy' breakthrough. https://www.theverge.com/2019/10/23/20928294/google-quantum-supremacy-sycamore-computer-qubit-milestone

Rasool RU, Ahmad HF, Rafique W, Qayyum A and Qadir J (2021) Quantum computing for healthcare: a review. TechRxiv. https://doi.org/10.36227/techrxiv.17198702.v2

Seabaugh A (2013) The tunneling transistor. IEEE Spectr. https://spectrum.ieee.org/semiconductors/devices/thetunneling-transistor 50:35

Sengupta K, Srivastava PR (2021) Quantum algorithm for quicker clinical prognostic analysis: an application and experimental study using CT scan images of COVID-19 patients. BMC Med Inform Decis Mak 21:227. https://doi.org/10.1186/s12911-021-01588-6

Swan M, dos Santos RP, Witte F (2022) Quantum Neurobiology. Quantum Rep 4:107–126

Chapter 24
E-Agriculture Diaspora: Heralding A New Era of Animal Farming and Agricultural Practices

Y. S. Jadoun, Chandra Sekhar Mukhopadhyay, Amandeep Singh, and Navkiran Kaur

Abstract The use of information and communication technologies (ICTs) in agriculture and the livestock sector production, research, and education has increased significantly. The ICT is primarily making a larger contribution to the production of livestock, the management of livestock herds, and the commercialization of livestock, agricultural products, and animal products. ICT is a tool that farmers may employ to deal with production productivity issues, use ecologically friendly and sustainable inputs and resources, and adapt crop practices and patterns to changing climate. The data acquired via ICT tools and applications can be used by authorities to target social programmes, gauge the availability of food, and change legislation. The technology may also be promoted by them. By incorporating ICT into national initiatives, encouraging an investment-friendly policy environment, and building consistent and interoperable digital systems, access for end users can be increased. Small-scale producers or farmers in developing countries may occasionally lack access to robust yet affordable technologies, but conducting impact studies and exchanging project data and knowledge helps focus and speed up the development of such ICTs in the agriculture and livestock sector.

Keywords E-Agriculture · Precise animal farming · Smart agricultural practices · Information and communication technology · Gadgets

Y. S. Jadoun (✉) · N. Kaur
Department of Veterinary & A.H. Extension Education, College of Veterinary Science, Guru Angad Dev Veterinary and Animal Sciences University (GADVASU), Ludhiana, Punjab, India
e-mail: yogendrasingh@gadvasu.in

C. S. Mukhopadhyay
College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India
e-mail: csmukhopadhyay@gadvasu.in

A. Singh
Directorate of Extension Education, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab, India

C. S. Mukhopadhyay et al. (eds.), *Biotechnological Interventions Augmenting Livestock Health and Production*, Livestock Diseases and Management,
https://doi.org/10.1007/978-981-99-2209-3_24

24.1 Introduction

All farmers, including small landowners, have something to gain from the potential revolution that information and communication technology (ICT) may bring to the Indian agricultural sector. The agricultural sector is the most important one since it provides employment for the great majority of rural residents in developing nations. ICTs can be used to effectively address the problems with traditional agriculture and play a vital role in enhancing the livelihood security of subsistence farmers. ICT contributes to the increase in demand for novel methods. By giving them better access to resources, improved agricultural and livestock technologies, efficient production policies, markets, banking, financial services, etc., it also assists in empowering the population. The use of information and communication technology (ICT) and digital solutions to support and profit from agriculture is known as "e-Agriculture."

The main goal of the project is to increase the profits for small-scale and marginal farmers around the world. In the "e-Agriculture community," which is a prerequisite on a global scale, people from all over the world exchange knowledge, concepts, and resources related to the use of information and communication technology (ICT) for upliftment of the rural masses and agricultural production. The rural population can improve their socioeconomic status and rural way of life through e-agriculture. The e-agriculture group aims to connect outlying and farmers having less land holding to institutions, networks, and data using digital tools. E-agriculture techniques encourage organic agriculture and guarantee environmental preservation. When we examine how e-agricultural operates, it becomes clear that the use of digital technology is encouraged among the stakeholders to support agricultural production in both rural and urban populations, making the sale of produce simple and intelligent. As a result, farmers in rural areas have immediate access to information. By having more knowledge, they can solve problems and generate possibilities in agriculture. E-agriculture aims to make use of better information and communication technologies and procedures to ensure that rural and agricultural growth is improved.

24.2 History of e-Agriculture

The main goal of WSIS was to lessen the disparity in access to ICT and other digital solutions that exists between rural and urban people worldwide. After that, AI-driven technologies are revolutionizing the cattle and agricultural industries. The inequity is a result of a pervasive issue brought on by ignorance, incorrect information exchange, a lack of proper planning to utilize the available electronic resources, etc. The e-Agriculture Community debuted to the general public in 2007. Users can keep up with the FAO e-regularly agriculture's scheduled activities at https://www.fao.org/e-agriculture/activity (FAO 2019).

24.3 Smart Technologies and ICT

The "Digital-India-Initiative" and telecom companies' ability to offer affordable rates to subscribers are two initiatives that the Indian government has fervently supported and encouraged. Additionally, the farmers can communicate with their peers via social media and cell phone services like Kisan Call Center (KCC). Together, these technologies create a "Information Web" for farmers and are in charge of disseminating up-to-date information for the advancement of agriculture and animals.

Resources for livestock are crucial for a developing nation like India. India must integrate animal husbandry with advancements in other industries if it wants to be a global leader in this sector. In both industrialized and developing nations, ICT has displayed considerable promise for revolutionizing extension services. One of the important areas that has the potential to alter the livestock industry is enhancing rural livelihoods using smart technologies. Information and communication technology (ICT) is a broad word that includes all communication technologies, including the internet, wireless networks, digital television, cell phones, satellite communications, etc. These are frequently referred to as "smart technologies" informally. ICT often refers to communications and computer-based technologies. A revolution in ICT research and development has resulted from the rapid expansion of mobile and internet users. If we referred to this era as the "Era of ICTs," we wouldn't be mistaken.

The employment of ICT technologies has the potential to alter India's agricultural, pastoral, and rural craft economies (Sasidhar and Sharma 2006). According to Tiwari and his colleagues (2010), the livestock industry should provide need-based, regionally relevant, and local language content for computer software and other electronic materials on dairy herd management, livestock production, marketing of livestock, and livestock disease control. Internet technology is now available to everyone owing to the Digital India plan of the Indian government and telecom companies' willingness to offer people affordable rates. Many of these ICTs have been created for use by livestock producers by various governmental and private entities engaged in research relevant to the livestock sector. RFID tags are being used to identify the animals, which helps livestock farmers as well as resource-based businesses with resource disposal. Additionally, social media and cell phone services like Kisan Call Center connect farmers to their peers.

Various applications of the new buzzword Artificial Intelligence (AI) have the potential to revolutionize the livestock industry. It is also claimed that AI will usher in a digital transformation known as the "fourth industrial revolution" on the world. The current proportion of AI in India's agriculture and allied sectors is barely 5.0%, but by 2030, it is predicted to double, according to NASSCOM (2017). There are numerous possibilities in which AI can be used by farmers, including the creation of learning simulations for those who want to switch from crop to livestock farming, the development of AI-based livestock expert systems, the development of algorithms to determine animal production, the understanding of patterns and numbers of

losses due to animal diseases and mortality, etc. Together, these technologies provide a "Information Web" for farmers and are in charge of disseminating timely information for the growth of cattle. The function of ICT in practices for managing and producing animals is described in this chapter.

24.4 ICT Applications

The majority of industries, including e-commerce, e-governance, finance, agriculture, education, medical, defense, and transportation, employ information and communications technology (ICT). Since radio and television broadcasts of livestock-related programmes, ICTs have been used in the cattle industry.

24.4.1 Radio

In India, All India Radio has always been a leader in radio research and development, especially among farmers. Radio has always had a strong place in the process of social transformation, largely due to its inherent devotion and, at best, ability to keep up with the times. Even though All India Radio has always been dedicated to supporting the farmers of the country, on February 15, 2004, it launched a new project termed narrowcasting in order to change the program's focus from agriculture to farmers or Kisan Vani.

24.4.2 Television

Television paved the way for extension initiatives and improved the effectiveness of information transfer. Since many years, livestock producers relied heavily on television to get the information they needed. In light of this, the government introduced DD Kisan, a 24-h television station devoted to Indian agriculture. It debuted on May 26, 2015. India's first channel devoted to farmers is called DD Kisan. It provides farmers with up-to-date weather and market information so they can plan accordingly. In order to best serve its intended audience, the 24-h Kisan Channel broadcasts up-to-date information on agriculture and related subjects. The channel provides information on cutting-edge production techniques used around the globe as well as global agribusiness R&D. IMD, IARI, Agricultural Universities, Krishi Vigyan Kendras (KVKs), etc. are partners with the channel.

24.4.3 Community Radio Stations (CRS)

In addition to commercial and public broadcasting, community radio offers a third type of radio transmission. Geographical and interest communities are served by community radio stations. They air material that is well-liked and pertinent by a local, niche audience but frequently ignored by commercial or mass-media broadcasters. The communities they serve run, own, and have influence over community radio stations. New Community Radio Guidelines were put into effect by the Indian government on November 16, 2006.

24.4.4 Radio Frequency Identification Device (RFID) Technology

Radio frequency identification is known as RFID. An antenna and integrated circuit are found on an RFID tag, and they are used to transmit data to an RFID reader, which transforms the radio waves into usable data. Through a communications link, this data is sent to the main computer system, where it is stored in a database and then processed for analysis.

RFID has been used in various areas of animal sciences like in Chitale Dairy Plant started in August 1996 at Pune, Maharashtra, ITGI (IFFCOTokio) started in August 2009 operating at Punjab, Gujarat, and Maharashtra, General Insurance Co. Ltd., Pashudhan Bima functional at Rajasthan and Odisha, NDRI, Karnal started in April 2008 at NDRI, Karnal, Lakshya Dairy, Jind, Haryana, initiated in November 2010 at Jind, Haryana, Sangamner Milk Union, Maharashtra started in 2011 at Maharashtra, Kopordem Farm at Valpoi, Sattari, North Goa initiated in 1996 at Goa, Gauseva and Gauchar Vikas Board (GGVB), Gujarat started in August 2017 at Gujarat.

24.4.5 Mobile Telephony

Cellular communication uses a technology called mobile technology. Over the past few years, mobile technology has advanced quickly. A typical mobile device has transformed since the start of this century from being nothing more than a simple two-way pager to incorporating a mobile phone, GPS navigational tool, an embedded web browser, an online messaging client, as well as a portable gaming console. There were 4.66 billion people using the internet actively as of January 2021 (59.5% of the global population). About 4.32 billion people, or 92.6% of this total, used mobile devices to access the internet. Mobile phones are used for making and receiving calls, sending SMS messages, accessing a variety of programmes, and interacting with social media. Few essential mobile applications used in animal husbandry venture are m-Kisan launched on July 2013 by NIC, GoI, Kisan Suvidha

started on March 2016 by the Department of Agriculture, GoI, Shetkari Masik App launched on 2014 developed by the Department of Agriculture, Maharashtra, Agri Market App started in 2015 by the Ministry of Agriculture and Farmers Welfare, m-KRISHI® AQUA launched on 2015 by Tata Consultancy Services—Innovation Lab, IVRI-Pashu Prajanan App started in 2017 by ICAR-IVRI, IVRIShookar Palan App launched in 2018 by ICAR-IVRI, IVRIArtificial Insemination App launched in 2018 by ICAR-IVRI, IVRI-Waste Management Guide App launched in 2019 by ICAR-IVRI, Information Network for Animal Productivity & Health (INAPH) developed by NDDB, Pashu Poshan App launched in 2015 by NDDB, NDDB AGR launched in 2016 by NDDB, GADVASU Precision Dairy Farming App started in 2017 by Guru Angad Dev Veterinary and Animal Sciences University (GADVASU), GADVASU Pig Farming App launched in 2020 by GADVASU, GADVASU—Goat Farming App launched in 2018 by GADVASU, GADVASU Dairy Prajnan App initiated in 2018 by GADVASU, GADVASU Antimicrobial Resistance App launched in 2021 by GADVASU, YODHA (Beta version) developed by GADVASU, PPTAK App launched in September 2022 by GADVASU and GADVASU Services App developed by GADVASU in September 2022.

24.4.6 Social Media

Approximately 50.6% of people worldwide used social media as of April 2021. More over 50% of Indians had access to social networks in 2020. According to the data, there were 2.85 billion active users of Facebook, followed by 2 billion active users of WhatsApp, 1.3 billion active users of Instagram, and 396 million active users of Twitter.

24.4.7 SMS

On July 16, 2013, the SMS portal known as m-Kisan was introduced. Farmers will receive SMS messages with information, services, and advisories. The "NDRI Messaging Portal," or Messaging-cum-IVRS for dairy farmers, was introduced in 2014 in six of the nation's key milk-producing states by ICAR-NDRI.

24.4.8 Kisan Call Centre

It is launched by the Indian government's ministry of agriculture. For Kisan Call Centers, a standard 11-digit toll-free number 1800-180-1551 or 1551 has been allocated. The number can be dialed in both landlines and mobile phones. The farmers receive responses in 22 regional languages to their questions. The

agrigraduates offered the farmer's initial response when they called the helpline's toll-free number. The call is transferred to a call centre representative if the farmers' questions answered by the agrigraduates are unsatisfactory or if the farmers need more information. There are IVRS-based responses available after hours.

24.4.9 Pashu Palak Tele-Advisory Kendra (PP-TAK) of GADVASU

PPTAK is a GADVASU, Ludhiana-based project funded by NABARD. It is a ground-breaking programme for farmers who raise livestock. The farming community and the university will forge a solid connection through communication, covering the entire state of Punjab. The university's call centre service will encourage livestock husbandry in places with a lack of information.

24.4.10 E-Vet Project

Major objective of the e-VET project is to make use of ICT in order to deliver veterinary services to livestock farmers in rural areas right at their doorsteps. Services offered include online advice and referrals for disorders and diseases affecting livestock, details on the costs of animals and livestock products at various marketplaces, and the creation of reports for monitoring needs.

The trained AVFOs and Gau-sevaks will be carrying a netbook and a cell phone with a camera when they travel to the selected villages to collect information and convey livestock medical problems and diseases to veterinarians at the block and district level veterinary institutions. These clinics' veterinarians review the provided information and upload the prescription electronically. Additionally, there will be an option to upload pictures of sick or injured animals, which will help the veterinarian comprehend the condition better. A brief video clip of the animal's condition can be posted in complicated circumstances for convenient reference.

24.4.11 Pashu Aadhaar: A Scheme to Digitize Animal Health

In August 2019, the Pashu Aadhaar system was launched to cover India's 94 million strong population of cows and buffaloes that are "in milk." Then, it was enlarged to encompass all cattle, including males, calves, elderly cattle, etc. Similar to Aadhaar cards, Pashu Aadhaar would likewise give each animal a 12-digit unique identification number, as well as the owner's information. Information of the pedigree of the cattle as well as details on calving, milk output, and immunization would all be

included in Pashu Aadhaar. It will improve the traceability of animals and assist farmers and officials in animal husbandry in coming up with better plans for managing livestock in the nation.

24.4.12 Goat Management

A free programme named Majordomo was used to launch an email conferencing system for goats at the Central Institute for Research on Goats (CIRG), Makhdoom (Mathura). The institute has established three email conferencing apparatus to facilitate crucial data between technicians, farmers, development officers, and planners.

24.5 Other ICT Tools Connecting Farmers

To communicate with the farmers, a variety of platforms are employed, including teleconferencing, videoconferencing, satellite phones, and emails.

24.5.1 Expert System

Expert systems are virtually identical to information systems; however, they are created for specific tasks like sickness prediction or milk fat analysis. Expert systems are more result-oriented than information systems; hence, their programming is more rigorous.

Examples of expert systems used in animal husbandry and veterinary practices are Automatic Milk Collection Unit Systems (AMCUS) started in1996 by the National Dairy Development Board operational in Gujarat, and in this system, the amount of milk, the amount of fat in the milk, and the amount owed to the farmer are all accurately displayed. Testing the milk quality right after as opposed to waiting 2–3 h after collection, the National Animal Disease Referral Expert System (NADRES) initiated in 2011 by NIVEDI for the whole country and based on historical incidence patterns, AR-NIVEDI has identified 13 priority diseases and amassed a substantial database on these illnesses. The database, which is the foundation of NADRES, is used to provide monthly livestock disease forewarning. These bulletins are created to inform animal husbandry departments at the national and state levels to implement the necessary control measures.

24.5.2 Web Portals and Websites

The term "Web portal" entered the Web IT lexicon in the late 1990s. A Web portal is a specifically created website that gathers data from several sources, such as emails, online forums, and search engines, in a consistent manner. While the website only has one-way communication, the web portal offers two-way conversation. Various other web-portals and their benefits are i-Kisan started in 1999 by Nagarjuna Group of Companies in Andhra Pradesh, the agricultural portal i-Kisan serves as a farmer's one-stop information hub. Online, in-depth information about crops, crop management methods, fertilizers, insecticides, and a wide range of other agriculture-related topics are available at i-Kisan. It also includes details about soils and feed. Electronic Knowledge Network for Innovators "Know Net-Grin initiated in 2000 operated in all parts of the country started in Gujarat, and it is a village-based kiosks with multimedia/multilingual databases connecting an electronic network of innovators with users and peers and communicating in local tongue. TARAhaat TARA was formed in 1983 by Technology and Action for Rural Advancement (TARA) operational in Punjab, Uttar Pradesh; its major benefit is establishment of sustainable livelihoods, particularly through the development and use of relevant technologies, and the dissemination of those technologies through microbusinesses in rural India. The creation of TARAhaat, which also focuses on livestock development, is Tarahaat.com.

Agropedia was started in January 2009 by the Government of India with assistance from the World Bank in all the states of the country; agropedia is an online resource for knowledge on India's agricultural industry and associated fields. With appropriate interfaces built in a mutual mode and in a variety of languages, it includes universal models and restricted material for a variety of users. It offers details on more effective fish farming and animal husbandry techniques. E Pashuhaat initiated in 2016 by the Department of Agriculture, GoI for the whole country, to make it convenient for there is no need to log in to use the portal in order to check the information. However, previous registration is required in order to complete any transaction. In Kisan started in 2013 by C-DAC, NIC for the whole country, to register farmers for the same, this portal and the m Kisan SMS service were launched. Farmers must register at this portal in order to subscribe to the m Kisan SMS service. Additionally, it registers farmers for K sewa, KCC, IVRS, and USSD. This webpage also includes a link to download apps for India's agricultural and associated industries. e-Pashu Palan Portal launched in 2020 by GADVASU operational in Punjab, GADVASU created this website with livestock producers in mind. Farmers can access this site to receive the bimonthly newsletter, numerous articles about livestock management techniques in the regional language, training forms for various training programmes, and frequent advising services.

Software for education created by ICAR-IVRI is information system for farmers in Hindi on "Pashudhan- avumKukkut Rog Suchna Pranali" (PAKRSP), this programme was created for the 78 most significant cattle and poultry diseases that are frequent in the nation as well as nine popular practice packages. With audio

backup, animations, and photos and line diagrams, the system is available in Hindi and would serve the entire nation. The CD will be particularly helpful for farmers raising cattle and poultry since it will enable them to treat most illnesses with first aid and will aid in the detection, prevention, and management of diseases. Goat Health Management Information System (GHMIS), it is a multi-language software that is available from the university and is available in Bengali, English, and Hindi. The owners of goats can learn about the numerous diseases that affect them, the symptoms that healthy and ill animals exhibit, the goats' vaccination and deworming schedules, and more. The technology is supported with language-specific voice, text, and high-quality images to effectively communicate information on goat health.

24.5.3 Educational CDs

The "Health Information System (HIS)" is a CD for farmers in Marathi which contains comprehensive knowledge on significant dairy animal diseases for which dairy owners need information. The Guru Angad Dev Veterinary and Animal Sciences University (GADVASU), in Ludhiana, Punjab, creates educational CDs on bovine reproduction.

24.5.4 Massive Open Online Course (MOOC)

An online course with open access and limitless enrollment is known as a massive open online course (MOOC). For the scientific community, there are a lot of online courses accessible linked to raising livestock, but there aren't many available for farmers. Currently, GADVASU, Ludhiana, is creating a MOOC on pig farming in the Hindi language. For veterinarians, the National Institute of Agricultural Extension Management (MANAGE) in Hyderabad offers a variety of MOOCs, including PGDAEM and others. The companies Coursera and Udemy are those that offer a variety of online, need-based MOOCs.

24.6 Artificial Intelligence and its Application in Livestock Sector

In India's agriculture and allied sectors, AI now accounts for only 5% of the total, but it is expected to quadruple by 2030. AI can be used in a variety of ways for livestock farmers, including the development of handy tools for disease identification, estimation of milk production, development of AI-based information and expert systems, estimation of disease losses, development of learning simulators, etc. Robots

cannot wipe out humanity, despite the fact that AI has both benefits and drawbacks. The ability to think imaginatively is something that only humans have; machines will never have it (Singh et al. 2021). Examples of use of AI in livestock sector are listed below.

24.6.1 Application in Livestock Health

S. No.	Application	Developed by	Utility
1	Livestock disease control	'National Disease Control Information System' (NDCIS) of New Zealand	Provides a database on important animal diseases like tuberculosis and brucellosis
2	Program for monitoring emerging diseases (ProMED)	International Society for Infectious Diseases (ISID) programme	Open to all sources, the global electronic reporting system for outbreaks of promising infectious illnesses and poisons
4	Robotic imaging	Penn State University's veterinary college	High-quality images are obtained
5	Canine patient simulator	Cornell's College of Veterinary Medicine	These pet simulators are cutting edge learning tools which are used to teach students. These are in line with animal ethics and welfare. Students can learn using these tools effectively without causing harm to the real animal Fletcher et al. (2012)
6	Thermal imaging cameras		These cameras offer a dependable non-contact technology that may be deployed rapidly. There is almost no need to medicate or touch animals, and they are almost never exposed to dangerous radiation. Singh et al. (2021)
7	Anti-stress ear tag for cattle		With robust, real-time animal state monitoring, the anti-stress ear tag enhances herd-wide production providing an analyses of about 200 physiological parameters. It helps in heat sensing and advice for the time of insemination
8	Pig respiratory disease package		To diagnose abnormal pig respiratory sounds which will further help in ruling out the disease

24.6.2 Applications in Livestock Production

S. No.	Application	Developed by	Utility
1	3D cameras to assess beef cattle		3D cameras has the potential in enhancing the livestock productivity. Generally, these cameras are used to assess beef cattle. These cameras take multiple pictures which are tested to form a convolutional neural network-based algorithm
2	Automatic feed manager		Automatic Feed Manager is a complex system based on sensors and predictive data analytics. Sensors identify any changes in the batch of the feed manufactured and alerts the manufacturer Karn et al. (2019)
3	Robo-cams for poultry	University of Georgia's (UGA)	The utilization of robot cams are feasible in poultry house; however, the studies are underway for their automation in poultry houses Poultry Tech (2016)
4	Virtual fences for controlling cattle	Marsh (1999)	The use of GPS technology to track the whereabouts of wildlife is common. Marsh's work to include bilateral stimulation, using separate sound stimuli for each ear, has led to the better control of animal
5	The Dutch cattle expert system (veePRO)	Dutch organisation named Veepro	This expert system may prescribe feed diets, treatments, and livestock health and welfare conditions. It also aids in animal reproduction by suggesting the mating partners whose progeny can lead to better production results

24.6.3 Applications for Animal Reproduction

S. No.	Application	Developed by	Utility
1	Smart neck collar		Smart collars have shown to be beneficial not only to health management but also to fertility. The smart neck collars are sensor-based equipment for recording various physiological parameters
2	Face recognition systems		Face recognition works on the principle of image analysis. The images pertaining to pattern of spots on animal body along with their actual face are analysed for generating results. It takes a few seconds for the system to distinguish a certain animal

(continued)

S. No.	Application	Developed by	Utility
3	Cow gait analyzer or Pedometry		Pedometry is based on the number of steps an animal walks a day. During estrus, the animals show restlessness and walk more footsteps. Analyses of the footsteps to ascertain the estrus behaviour of animal is pedometry
4	Intelligent dairy assistant		Intelligent dairy assistant serves as an aide to the livestock farmers for management of their dairy animals. It was developed by a Dutch company to track the movements of the dairy animals
5	MSUES cattle calculator	Mississippi State University's extension	It is beneficial for the users rearing beef cattle. A reproductive calculator is included in the software for the calculation of breeding and calving periods

24.6.4 Application for Livestock Products

S. No.	Application	Developed by	Utility
1	Robotic milking systems or automatic milking systems (AMS)		The AMS have been developed for reducing the time management constraint in dairy operations. This is based on voluntary milking principle whereby a dairy animal decide the time of milking and interval between milking on its own
2	Robotic hide puller		Robotic hide puller signifies an era of automation in meat industry. It is based on increasing automation and reducing human touch. This is basically an instrument used to remove animal hides after slaughter
3	Smart packaging		AI-based packaging is replacing the conventional laser-based packaging. Cortex system is an AI-based system for livestock products packaging which consists of a camera with computer vision. The camera scans the products passing through the conveyor belt and removes the faulty ones from the production line
4	E-nose or E-tongue		Electronic nose or electronic tongue comes under the ambit of electronic sensing or e-sensing. It is a set of gas or chemical sensors that are embedded in an instrument and work together to form a complete sense of taste, smell

(continued)

S. No.	Application	Developed by	Utility
			and flavour. 'Electronic nose (e-nose)' consists of gas sensor arrays, whereas 'electronic tongue (e-tongue)' consists of chemical sensor arrays
5	Meat quality evaluation using computer vision		Computer vision (CV) or imaging technology has gotten a lot of attention as a non-destructive and quick way to measure the quality aspects of agricultural products, including meat and meat products, all over the world
6	Bio-sensing technology		Biosensors are novel aids to ensure food safety which works on the principle of conversion of chemical signals into electronic signals and thereby detecting a hazard in food product
7	AI based meat sorter		It has a special place in quality control of the product. The product which does not meet the quality requirements are sorted in the initial stages which imparts consumer preference to the product
8	CNN-based meat identification		The Convolutional Neural Network (CNN) is a deep learning technique which is used frequently in the classification type of inputs. CNN learns from the input images and trains itself from a large dataset Krizhevsky et al. (2017)
9	Ascertaining carcass quality or classification		This human involvement to ensure carcass quality can be reduced by using convolution neural networks which will ensure quality along with hygiene in the slaughterhouses. Furthermore, computer vision can also be used for this purpose
10	AI based cameras for food safety compliance		AI-based cameras can be used at eateries, processing plants, restaurants to ensure meat safety and hygiene. These cameras can detect whether the employees are wearing proper safety suits along with tracking their movement and physiological status
11	Intelligent cleaning systems		If we compare, traditional cleaning methods with the intelligent cleaning systems, it can be found that former cannot remove minute food particles which leads to pathogen build-up thus reducing the quality of the product
12	Development of meat products		Machine learning algorithms used by AI play a crucial part in precisely

(continued)

S. No.	Application	Developed by	Utility
			adding ingredients to a meat product, as well as managing temperature and processing conditions
13	Meat supply chain optimization		Neural network-based algorithms can calculate the present supply and future demand of the meat products. The marketing of products can be optimized by monitoring the demand and supply of products
14	Marketing of Livestock Products	National Dairy Development Board	This method has reduced the alteration of milk and prompted payment to the farmers. It is because of transparency in milk marketing, the dairy sector in India has seen an unparalleled growth Sharma (2000)

24.6.5 Applications for Animal Welfare

S. No.	Application	Developed by	Utility
1	Robot fish	Massachusetts Institute of Technology in 1989	The improvement of robot fish control and navigation is the most significant area of their research and development, as it allows them to 'communicate' with their environment, allowing them to go along a certain course and respond to commands to make their 'fins' flap
2	Protection assistant for wildlife security (PAWS)		It is an integrated module to prevent poaching which fetches data from the previously poached areas and routes and predict regions and routes where future poaching can take place. This module is based on machine learning Lemieux (2014)
3	Man's best friend 2.0		This pet dog provides companionship along with utility for the family members
4	Minimizing drug testing on animals	Insilico Medicine	The systems produce good predictions without the use of animals if given enough data, although traditional testing is still required in some circumstances

24.6.6 Applications for Livestock Statistics

There are many AI-based software which are used for analyses of data and interpretation of results. From disease diagnosis to computational genomics, the AI-based software have entered animal sciences and proving their utility. Few of the software are enlisted below for reference.

- Vetel's Diagnostic Software
- IBM's Vet Computing Tool
- Sofie Cognitive Computing Tool
- Deep Mind for Record Keeping
- Deep Genomics

24.7 Epilogue

The country's significant rise in internet users demonstrates the prospective role that ICTs may play in the growth of livestock, provided that this potential is realised quickly and effectively. The implementation of need-based digital literacy programmes for livestock owners would help to address issues like digital illiteracy, which are still a barrier to technological interventions in the livestock sector. Another problem that can be resolved is insufficient telecom coverage by establishing cellular towers in outlying areas. Technology creators must also assess the information needs of the rural population and develop the technologies that will meet those needs. Building the capacities of stakeholders and livestock producers is necessary for the creation of ICTs and their use, respectively. Modern treatments like geographic information systems, remote sensing, and AI-based ICT tools have a bright future, but it is important to assess their cost-effectiveness and perceived utility in rural India. Sustainable production and environmental management can result from the use of ICTs for the benefit of the farming community. Making ICT tools more interactive, user-friendly, and economical should be the emphasis of future research.

References

FAO (2019). https://www.fao.org/e-agriculture

Fletcher DJ, Militello R, Schoeffler GL et al (2012) Development and evaluation of a high-fidelity canine patient simulator for veterinary clinical training. J Vet Med Educ 39(1):7–12. https://doi.org/10.3138/jvme.0711.073R

Karn P, Sitikhu P, Somai N (2019) Automatic cattle feeding system. In: 2nd International conference on engineering and technology; KEC Conference-2019 at Dhapakhel, Lalitpur, Nepal

Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. Commun ACM 60(6):84–90. https://doi.org/10.1145/3065386

Lemieux AM (2014) Situational prevention of poaching. Routledge, Crime Science Series

Poultry Tech (2016) Researchers develop robot to autonomously operate in a poultry Growout house. Poultry Tech 28(1):1–2

Sasidhar PVK, Sharma VP (2006) Cyber livestock outreach services in India: a model framework. Livest Res Rural Dev 18(2):1–12. http://www.lrrd.org/lrrd18/1/sasi18002.htm

Sharma VP (2000) Cyber extension in the context of agricultural extension in India. Man Exten Res Rev 1(1):24–41

Singh A, Brar PS, Mehta N et al (2021) Artificial intelligence in livestock sector: today and tomorrow. In: Brar PS, Mehta N, Singh A, Sivakumar S, Phand S (eds) Value addition of milk and meat: a push to entrepreneurship. Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana & National Institute of Agricultural Extension Management, Hyderabad, pp 131–144

Tiwari R, Phand S, Sharma MC (2010) Status and scope of information and communication technology for livestock and poultry production in India–a review. Ind J Animal Sci 80(12): 1235–1242